

DECENTRALIZED SEQUENTIAL CHANGE DETECTION*

BY GEORGIOS FELLOURIS AND GEORGE V. MOUSTAKIDES

University of Southern California and University of Patras

The problem of decentralized sequential change detection is considered, where an abrupt change occurs in an area that is being monitored by a number of sensors. The goal is to detect this change as soon as possible at a central location (fusion center) which receives information from all sensors subject to quantization and rate constraints. A novel decentralized sequential detection rule is proposed that requires communication from the sensors at random times and transmission of only low-bit messages. The second-order asymptotic optimality of the proposed scheme is established under different statistical models for the sensor observations. Specifically, when each sensor process either has continuous paths and is continuously observed or it is a random walk, it is proved that the inflicted performance loss (with respect to the optimal detection rule that uses the complete sensor observations) is bounded asymptotically as the rate of false alarms goes to 0. The proposed scheme remains asymptotically optimal but of first-order even if it induces an asymptotically low communication rate and there is an asymptotically large number of sensors. Finally, simulation experiments illustrate its efficiency in practice and its superiority over alternative decentralized detection rules that rely on communication at deterministic times.

1. Introduction. Suppose that an area is being monitored by a number of sensors which transmit their observations to a central location (fusion center). At some unknown time, an abrupt “disorder” occurs in the monitored area, such as an unexpected intrusion, and changes the dynamics of all sensors. Assuming that the sensors acquire their observations sequentially, the goal is to raise an alarm at the fusion center, using the transmitted messages from all sensors, as soon as possible after the occurrence of the change.

When the sensors transmit their complete observations to the fusion center, this is the classical problem of sequential change detection, which has applications in many scientific and industrial fields, such as industrial quality

*This work was supported in part by the US National Science Foundation under Grant CIF1064575.

AMS 2000 subject classifications: Primary62L10, 60G40

Keywords and phrases: CUSUM, Change-point detection, CUSUM, Decentralized detection, Communication constraints, Quantization, Random sampling, Asymptotic optimality.

control [27], computer security [31] and many others. For the methodological and theoretical developments in sequential change detection we refer to the books by Baseville and Nikivorov [1]; Poor and Hadjiliadis [24]; as well as the review articles by Lai [10], [11]; Shiryaev [30]; Polunchenko and Tartakovsky [23].

In many modern application areas, such as mobile and wireless communications and distributed surveillance systems, the sensors are typically low-power devices with limited energy, whereas their links with the fusion center are characterized by limited communication bandwidth [25],[36]. Therefore, in order to preserve the robustness of the network, it is necessary to limit the overall communication load and in particular the communication activity of each sensor. This primarily implies a *quantization* constraint, i.e. each sensor should transmit a small number of bits each time it communicates with the fusion center, but also a *rate* constraint, i.e. each sensor should communicate with the fusion center at a lower rate than its sampling rate. In these cases, the problem at hand is first to decide the information that should be transmitted from the sensors to the fusion center, respecting the above constraints, and then to construct a sequential detection rule at the fusion center that relies on this information. We will call such detection rules *decentralized*, in contrast to the *centralized* ones that rely on the full sensor observations.

There is a number of articles that study decentralized sequential detection rules, for example we refer to Crow and Schwartz [4], Veeravalli [35], [36], Mei [14], Moustakides [18], Tartakovsky and Veeravalli [8], [32]. In most papers, the emphasis is placed on the quantization constraint and, typically, one-bit transmission is imposed. With respect to the rate of communication, two extreme cases have been considered in the literature. On the one hand, it is often assumed that each sensor transmits a quantized version of *every* observation it takes, i.e. the communication rate is equated with the sampling rate. On the other hand, one-shot schemes have been explored, which require each sensor to communicate with the fusion center *at most once* and to transmit a single bit of information.

However, even if two detection structure rules require the transmission of only one-bit per communication, they may actually induce very different transmission activities, if the one communicates at a high rate and the other rarely. Therefore, comparing them may be misleading, especially in a decentralized setup. In order to highlight this point, we formulate a general framework for the problem of decentralized sequential change detection, which encompasses most of the schemes that have been proposed in the literature.

The main contribution of this work is that we propose a novel decentralized detection rule and establish its asymptotic optimality under a large class of sensor dynamics. In particular, we suggest that each sensor communicate whenever its local log-likelihood ratio exits an interval and inform the fusion center with a *low-bit* message regarding the evolution of this statistic since the previous communication time. The fusion center then uses the transmitted messages in order to detect the change. Similar communication schemes have been used in the context of the decentralized sequential hypothesis testing problem by Fellouris and Moustakides [6] and Yilmaz et al. [38]. As a result, the communication activity of each sensor is completely controlled through the selection of three parameters, the upper and lower bounds of the interval and the number of bits the sensor transmits per communication.

When the local log-likelihood ratios have continuous paths or they are random walks, we show that there is only bounded performance loss with respect to the optimal centralized detection rule as the period of false alarms goes to infinity (second-order asymptotic optimality). Moreover, we show that the proposed scheme remains asymptotically optimal (of first-order) even when it induces an asymptotically low communication rate and there is an asymptotically large number of sensors. Finally, we illustrate with simulation experiments its superiority over a decentralized detection rule that requires communication at deterministic times.

The structure of the remaining paper is the following: in Section 2, we formulate the problems of centralized and decentralized sequential change detection. In Section 3, we describe the main decentralized schemes in the literature. In Sections 4 and 5, we define and analyze the proposed scheme in continuous and discrete time, respectively. In Section 6, we summarize our results and state our conclusions. We prove the main results of the paper, as well as some supporting lemmas, in Appendices A, B and C.

2. Sequential Change Detection. Let $\{(\xi_t := \xi_t^1, \dots, \xi_t^K)\}$ be a K -dimensional stochastic process, where $\xi_0^k := 0$ for every $1 \leq k \leq K$ and time is either discrete ($t \in \mathbb{N}$) or continuous ($t \in [0, \infty)$). The interpretation is that ξ^k is the process observed by sensor k . Thus, if $\{\mathcal{F}_t^k\}$ is the local filtration at sensor k and $\{\mathcal{F}_t\}$ the global filtration, then it is $\mathcal{F}_t^k := \sigma(\xi_s^k, 0 \leq s \leq t)$ and $\mathcal{F}_t := \vee_k \mathcal{F}_t^k$ for every t . Moreover, it is understood that we work with the right-continuous versions of these filtrations whenever time is continuous.

Let P_0 and P_∞ be two probability measures on the canonical space of ξ that are mutually absolutely continuous on any σ -algebra \mathcal{F}_t . We denote by

u_t the corresponding log-likelihood ratio up to time t , that is

$$(2.1) \quad u_t := \log \frac{dP_0}{dP_\infty} \Big|_{\mathcal{F}_t}.$$

We assume that at some unknown, deterministic time $\tau \geq 0$, the distribution of ξ , which we denote by P_τ , changes from P_∞ to P_0 . Therefore, P_τ coincides with P_∞ on \mathcal{F}_t when $t \in [0, \tau]$ and is absolutely continuous with respect to P_∞ on \mathcal{F}_t when $t > \tau$ so that

$$(2.2) \quad \log \frac{dP_\tau}{dP_\infty} \Big|_{\mathcal{F}_t} = u_t - u_\tau, \quad t > \tau.$$

2.1. The centralized setup. The problem of classical (centralized) change detection is to find an $\{\mathcal{F}_t\}$ -stopping time that has small detection delay and rare false alarms, i.e. \mathcal{T} should take large values under P_∞ and $\mathcal{T} - \tau$ should take small values under P_τ . However, finding the optimal detection rule depends on how detection delay and false alarms are quantified. There are different approaches to the sequential change detection problem, such as the Bayesian formulation due to Shiryaev [28] (see also Peskir and Shiryaev [21], Gapeev [7], Dayanik et al. [5] and Sezer [26]) and the minimax formulation due to Pollak [22] (see also Poluchenco and Tartakovsky [23] and Tartakovsky et al. [33]). In this work, we focus on the formulation suggested by Lorden [13], where the performance of a detection rule \mathcal{T} is measured by its worst-case (with respect to τ) conditional expected delay given the worst possible history of observations up to τ ,

$$(2.3) \quad \mathcal{J}_L[\mathcal{T}] = \sup_{\tau \geq 0} \text{ess sup } E_\tau \left[(\mathcal{T} - \tau)^+ \mid \mathcal{F}_\tau \right]$$

and the optimal detection rule was defined as the solution to the following constrained optimization problem

$$(2.4) \quad \inf_{\mathcal{T}} \mathcal{J}_L[\mathcal{T}] \text{ when } E_\infty[\mathcal{T}] \geq \gamma,$$

where γ is a positive constant, fixed in advance by the designer of the scheme. In other words, the goal in this strongly min-max approach is to minimize the detection delay under the worst-case scenario with respect to both the changepoint and the history of observations before the change, while controlling the period of false alarms above a desired level, γ .

A related formulation was considered by Moustakides in [17], according to which the performance measure \mathcal{J}_L is replaced by

$$(2.5) \quad \mathcal{J}_M[\mathcal{T}] := \sup_{\tau \geq 0} \text{ess sup } E_\tau \left[\left(\langle u \rangle_{\mathcal{T}} - \langle u \rangle_\tau \right)^+ \mid \mathcal{F}_\tau \right]$$

and the optimal detection rule is defined as the solution to the following optimization problem

$$(2.6) \quad \inf_{\mathcal{T}} \mathcal{J}_M[\mathcal{T}] \text{ when } \mathbb{E}_\infty[\langle u \rangle_{\mathcal{T}}] \geq \gamma,$$

where $\langle u \rangle_t$ is the quadratic variation of the log-likelihood ratio at time t . Thus, in this formulation, detection delay and false alarm rate are not measured in terms of the actual time, but in terms of the expected accumulated quadratic variation until the alarm. The latter criterion also has an appealing interpretation in terms of Kullback-Leibler divergence, since $\mathbb{E}_0[u_{\mathcal{T}}] = \mathbb{E}_0[\langle u \rangle_{\mathcal{T}}]$ and $\mathbb{E}_\infty[-u_{\mathcal{T}}] = \mathbb{E}_\infty[\langle u \rangle_{\mathcal{T}}]$, whenever these quantities are finite. However, the main advantage of the latter formulation is that it admits a solution for much richer class of dynamics.

In order to be more precise, let us present the Cumulative Sums (CUSUM) test, which was introduced by Page [20] and can be defined as follows:

$$(2.7) \quad \mathcal{S} := \inf\{t \geq 0 : y_t \geq \nu\}, \text{ where } y_t := u_t - \inf_{0 \leq s < t} u_s,$$

and $\nu > 0$ is a fixed threshold. When $\{u_t\}_{t \in \mathbb{N}}$ is a random walk, it is well-known (see Moustakides [15], [16]) that the CUSUM test solves Lorden's [13] optimization problem (2.4), as long as ν is chosen so that the false alarm constraint be satisfied with equality, that is $\mathbb{E}_\infty[\mathcal{S}] = \gamma$. In continuous-time, when the process $\{u_t\}_{t \geq 0}$ has continuous paths and the following condition is satisfied

$$(2.8) \quad \lim_{t \rightarrow \infty} \langle u \rangle_t = \infty \quad \mathbb{P}_0, \mathbb{P}_\infty - \text{a.s.},$$

the CUSUM test solves (2.6), as long as its threshold ν is now chosen so that $\mathbb{E}_\infty[\langle u \rangle_{\mathcal{S}}] = \gamma$ (see Moustakides [17] and Chronopoulou and Fellouris [3]). A direct consequence of the latter optimality result is that the CUSUM test is also optimal with respect to Lorden's original criterion (2.4) when u_t has continuous paths and $\langle u \rangle_t$ is proportional to t . This is the case for example when each ξ^k is a fractional Brownian motion (fBm) with Hurst index H before the change and adopts a polynomial drift term with exponent $H + 1/2$ after the change. In the special case $H = 1/2$, this implies the optimality of the CUSUM test when each ξ^k is a Brownian motion that adopts a linear drift after the change, which had originally been shown by Shiryaev [29] and Beibel [2].

In what follows, in order to work with a common criterion, we quantify the performance of a detection rule T with

$$(2.9) \quad \mathcal{J}[T] := \sup_{\tau \geq 0} \text{ess sup } \mathbb{E}_\tau \left[(u_T - u_\tau) \mathbb{1}_{\{T > \tau\}} \mid \mathcal{F}_\tau \right]$$

and we define the optimal detection rule as the solution to the following optimization problem

$$(2.10) \quad \inf_{\mathcal{T}} \mathcal{J}[\mathcal{T}] \text{ when } \mathbb{E}_{\infty}[-u_{\mathcal{T}}] \geq \gamma.$$

When $\{u_t\}$ is a random walk, this problem is equivalent to Lorden's optimization problem, defined by (2.3)-(2.4), as long as we restrict ourselves to integrable stopping times under $\mathbb{P}_0, \mathbb{P}_{\infty}$. Similarly, when $\{u_t\}$ has continuous paths, this problem is equivalent to the modified version of Lorden's criterion, defined by (2.5)-(2.6), as long as we consider stopping times that satisfy $\mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{T}}] < \infty$ and $\mathbb{E}_0[\langle u \rangle_{\mathcal{T}}] < \infty$. Therefore, under all dynamics for which it is known to have an exact optimality property, the CUSUM test solves the problem defined by (2.9)-(2.10), given that its threshold is chosen so that $\mathbb{E}_{\infty}[-u_S] = \gamma$, which will be our standing assumption from now on.

2.2. The decentralized setup. In a decentralized setup, the goal is first to select a *communication scheme* subject to quantization and rate constraints and then to find a detection rule that is adapted to the filtration that is induced by the chosen communication scheme.

More specifically, we define a *decentralized* sequential detection rule as a pair $(\{\tilde{\mathcal{F}}_t\}, \mathcal{T})$, where \mathcal{T} is an $\{\tilde{\mathcal{F}}_t\}$ -stopping time and $\{\tilde{\mathcal{F}}_t\}$ is a filtration of the form

$$(2.11) \quad \tilde{\mathcal{F}}_t := \sigma((\tau_n^k, z_n^k) : \tau_n^k \leq t, k = 1, \dots, K),$$

where each $\{\tau_n^k\}_{n \in \mathbb{N}}$ is the sequence of communication times for sensor k and z_n^k is the message transmitted to the fusion center at time τ_n^k . Each τ_n^k must be an $\{\mathcal{F}_t^k\}$ -stopping time and each z_n^k an $\mathcal{F}_{\tau_n^k}^k$ -measurable random variable that takes values in a *finite* set, so that a small number of bits is required for its transmission to the fusion center. Moreover, since many applications are characterized by limited storage capacity, we require in particular that each z_n^k is measurable with respect to the σ -algebra generated by the observations at sensor k between its $n-1$ and n th transmission, that is $\sigma(\xi_s^k, \tau_{n-1}^k \leq s \leq \tau_n^k)$.

Note that this framework allows only one-way communication from the sensors to the fusion center, thus it forbids any communication between sensors or feedback from the fusion center to the sensors. Indeed, such possibilities impose a much heavier communication load in the network and raise questions regarding the design of the network architecture, which we do not consider here. Furthermore, under the assumption of independence across sensors, it is intuitive that such possibilities should be redundant and

one of the goals of this work is to show that this is indeed the case. For decentralized detection rules that require feedback we refer to Veeravalli [35].

2.3. Asymptotic optimality. Ideally, we would like to find the best decentralized detection rule, optimizing with respect to both the fusion center filtration and stopping time. Such an optimization problem is highly intractable, even if one makes a number of simplifying assumptions [35]. For this reason, we will use the centralized CUSUM as the ultimate benchmark and compare any decentralized detection rule against it. We can only hope that such a detection rule attains the optimal centralized performance asymptotically, that is for large periods of false alarms.

Thus, in what follows, if $(\{\mathcal{F}_t\}, \mathcal{T})$ is an arbitrary *decentralized* detection rule satisfying the false alarm constraint $-\mathbb{E}_\infty[u_{\mathcal{T}}] \geq \gamma$ and \mathcal{S} is the *centralized CUSUM rule* satisfying the false alarm constraint with equality, i.e. $-\mathbb{E}_\infty[u_{\mathcal{S}}] = \gamma$, we will say that \mathcal{T} is asymptotically optimal

- of order-1, if $\mathcal{J}[\mathcal{T}]/\mathcal{J}[\mathcal{S}] \rightarrow 1$,
- of order-2, if $\mathcal{J}[\mathcal{T}] - \mathcal{J}[\mathcal{S}] = \mathcal{O}(1)$,
- of order-3, if $\mathcal{J}[\mathcal{T}] - \mathcal{J}[\mathcal{S}] = o(1)$,

as $\gamma \rightarrow \infty$. Note that, contrary to order-1, order-2 asymptotic optimality guarantees that the performance loss of \mathcal{T} remains bounded as $\gamma \rightarrow \infty$. Of course, it is even better if the performance loss vanishes as $\gamma \rightarrow \infty$, which is the case of order-3 asymptotic optimality. Undoubtedly, order-3 implies order-2 which implies order-1, since $\mathcal{J}[\mathcal{T}], \mathcal{J}[\mathcal{S}] \rightarrow \infty$ as $\gamma \rightarrow \infty$.

3. Existing decentralized schemes. The main decentralized detection rules encountered in the literature can be classified into two main categories. In the first, the sensors transmit systematically compressed versions of their data to the fusion center and the latter combines these quantized messages in order to detect the change. In the second, each sensor detects individually the change and the fusion center combines the local sensor decisions.

In order to describe them in more detail, we need to introduce some additional notation and assumptions. Thus, we denote by \mathbf{P}_0^k and \mathbf{P}_∞^k the post and pre-change measure of ξ^k and by u_t^k their log-likelihood ratio up to time t , that is

$$(3.1) \quad u_t^k := \log \frac{d\mathbf{P}_0^k}{d\mathbf{P}_\infty^k} \Big|_{\mathcal{F}_t^k}.$$

Moreover, we assume that the following local and average Kullback-Leibler information numbers are positive and finite

$$(3.2) \quad I_0^k := \mathbb{E}_0[u_1^k], \quad I_\infty^k := -\mathbb{E}_\infty[u_1^k], \quad \bar{I}_0 := \frac{1}{K} \mathbb{E}_0[u_1], \quad \bar{I}_\infty := \frac{1}{K} \mathbb{E}_\infty[-u_1].$$

Furthermore, we assume that the sensor processes are independent, that is $\mathbf{P}_\infty := \mathbf{P}_\infty^1 \times \dots \times \mathbf{P}_\infty^K$ and $\mathbf{P}_\tau := \mathbf{P}_\tau^1 \times \dots \times \mathbf{P}_\tau^K$ for any τ , which implies

$$(3.3) \quad u_t := u_t^1 + \dots + u_t^K, \quad t \geq 0$$

and consequently $\bar{I}_0 = \frac{1}{K} \sum_{k=1}^K I_0^k$ and $\bar{I}_\infty = \frac{1}{K} \sum_{k=1}^K I_\infty^k$.

3.1. *Q-CUSUM.* Suppose that each sensor transmits to the fusion center quantized versions of its local log-likelihood ratio process at deterministic equidistant times. Thus, if r is the communication period and for each sensor k we consider the alphabet $\{1, \dots, b^k\}$, where $b^k \geq 2$ is an integer, then it will be

$$(3.4) \quad \tau_n^k = rn \text{ and } z_n^k = j \text{ when } \Gamma_{j-1}^k \leq u_{\tau_n^k}^k - u_{\tau_{n-1}^k}^k < \Gamma_j^k,$$

with $j = 1, \dots, b^k$, where $-\infty = \Gamma_0^k < \Gamma_1^k < \dots < \Gamma_{b^k}^k = \infty$ are fixed thresholds, chosen by the designer. When the sensors take continuous-time observations, the communication rate $1/r$ is clearly smaller than the (infinite) sampling rate (for any positive number r). On the other hand, when the sensors take discrete-time observations, the communication period is r times larger than the sampling period, where now $r \geq 1$ is an integer.

The communication scheme (3.4) induces synchronous communication to the fusion center, which receives at each time $\tau_n^k = rn$ the K -dimensional vector (z_n^1, \dots, z_n^K) . If we additionally assume that each $\{u_t^k\}$ has stationary and independent increments, then a natural detection rule at the fusion center is the corresponding CUSUM stopping time

$$(3.5) \quad \hat{\mathcal{S}} := r \cdot \inf\{n \in \mathbb{N} : \hat{y}_n \geq \hat{\nu}\},$$

where the threshold $\hat{\nu}$ is chosen so that the false alarm constraint is satisfied with equality and the CUSUM statistic $\{\hat{y}_n\}$ admits the following recursion:

$$(3.6) \quad \hat{y}_n := (\hat{y}_{n-1})^+ + \sum_{k=1}^K \sum_{j=1}^{b^k} \left[\mathbb{1}_{\{z_n^k=j\}} \log \frac{\mathbf{P}_0(z_n^k=j)}{\mathbf{P}_\infty(z_n^k=j)} \right], \quad \hat{y}_0 := 0,$$

One might claim that, *given the quantization rule (3.4)*, the CUSUM rule defined in (3.5)-(3.6) is the best we can do at the fusion center. We should

note however that this claim we know it to be true only when $r = 1$. When $r > 1$, the messages sent to the fusion center are not i.i.d. before and after the change, which is the crucial property for the optimality of CUSUM. Indeed, the change can take place *between* two consecutive communication instances, thus generating non i.i.d. data. If of course one could demonstrate that the worst-case scenario is the change to occur at a communication instant, then this would clearly establish optimality for the CUSUM test at the fusion center for *any* $r \geq 1$.

We call this detection scheme Q-CUSUM where Q stands for the “quantization” employed by this method. Note that we have to multiply by r in (3.5) in order to return to physical time units, since the samples are at a rate $1/r$. It is straightforward to see that as $\gamma \rightarrow \infty$

$$(3.7) \quad \frac{\mathcal{J}[\hat{\mathcal{S}}]}{\mathcal{J}[\mathcal{S}]} \rightarrow \frac{r\bar{I}_0}{\hat{I}_0}, \quad \hat{I}_0 := \frac{1}{K} \sum_{i=1}^K \sum_{j=1}^{b^k} P_0(z_n^k = j) \log \frac{P_0(z_n^k = j)}{P_\infty(z_n^k = j)},$$

where \bar{I}_0 is the average Kullback-Leibler information number defined in (3.2). Therefore, the asymptotic performance of $\hat{\mathcal{S}}$ can be optimized by choosing the thresholds $\{\Gamma_j^k\}$ in order to maximize \hat{I}_0 . However, it is well-known (see for example [34]) that for any choice of thresholds, communication period and alphabet size, it is $r\bar{I}_0 > \hat{I}_0$, therefore $\hat{\mathcal{S}}$ is not even asymptotically optimum of order-1.

In the case that the sensors take discrete-time observations, this detection rule has been studied in [4], [14], [18], [32] under the assumption that $r = 1$, i.e. when each sensor communicates with the fusion center at *every* observation time. In the case that the sensors observe Brownian motion paths in continuous-time, the performance of this rule was explored in [18].

3.2. Fusion of local CUSUM rules. We now consider the class of fusion center detection rules that rely on the local decisions of the sensors. More specifically, we assume that sensor k communicates at the following times

$$(3.8) \quad \tau_n^k = \inf\{t \geq \tau_{n-1}^k : y_t^k \geq c^k\},$$

where $y_t^k := u_t^k - \min_{0 \leq s \leq t} u_s^k$ is the local CUSUM statistic at sensor k and c^k is a positive threshold. With this communication scheme, a sensor transmits a message to the fusion center only to announce that it has detected a change and this requires only *one* bit per transmission. Thus, even if the network can support the transmission of *multi-bit* messages, this additional flexibility is not going to be useful for this communication scheme.

There are many reasonable fusion center policies that can be based on (3.8). For example, the fusion center may raise an alarm at the first time some sensor communicates, i.e. at $\min_k \tau_1^k$ (min-CUSUM). Alternatively, the alarm can be raised when every sensor has communicated at least once, i.e. at $\max_k \tau_1^k$ (max-CUSUM). It is clear that both schemes require minimal communication activity in the network, as they require the transmission of at most one bit from each sensor. The exact performance of these two schemes was computed in [18] in the case that each sensor observes a Brownian path. In the case of discrete-time i.i.d observations, it was shown in [32] that they both have the same first-order asymptotic performance. However, numerical experiments in both papers suggest that the min-CUSUM performs in practice much better than the max-CUSUM. Of course, as one would expect, both schemes are asymptotically suboptimal in all the above cases. (We should note however that the min-CUSUM is an appealing detection rule in the case that the change may occur in at most one sensor, see for example [8]).

An alternative detection rule that is based on the communication scheme (3.8) is to raise an alarm the first time that all sensors communicate *at the same time*, that is, at

$$(3.9) \quad \mathcal{M} := \inf\{t : y_t^k \geq c^k, \forall k = 1, \dots, K\}.$$

Thus, contrary to the previous *one-shot* schemes, each sensor now keeps transmitting messages to the fusion center, even after it has detected a change, until they all agree *simultaneously* that the change has indeed occurred. In this way, the induced communication activity is intense only after the change has occurred. Before the change, a sensor communicates only to report a local false alarm, which is a rare event.

This rule was suggested and analyzed by Mei¹ in [14], where it was shown that when each $\{u_t^k\}$ is a random walk whose increments have a finite second moment, \mathcal{M} is asymptotically optimal of first order, as long as each threshold c^k is chosen to be proportional to I_0^k (the constant of proportionality is determined by the false alarm constraint). This is the only asymptotically optimal decentralized detection rule that is known in the literature so far. However, \mathcal{M} is not asymptotically optimal of second order, since $\mathcal{J}[\mathcal{M}] - \mathcal{J}[\mathcal{S}] = \mathcal{O}(\sqrt{\log \gamma})$ as $\gamma \rightarrow \infty$ [14]. Moreover, despite its asymptotic optimality, \mathcal{M} can be inefficient in practice, especially in the case that K is

¹The presentation of \mathcal{M} is different in [14] where it is assumed that each sensor must communicate at every observation time t the outcome of the event $\{y_t^k \geq c^k\}$. However, it is easy to realize, that communication is needed only when the local statistic exceeds the local threshold.

large. These points were illustrated numerically in [14] and in [32]. Finally, we should note that, contrary to Q-CUSUM, \mathcal{M} does not have any degrees of freedom that allow the designer to control the induced communication activity, as the design parameters $\{c^k\}$ in (3.9) are completely specified by the desired period of false alarms.

3.3. Comparisons. The min-CUSUM and Mei's scheme have been compared numerically with Q-CUSUM in a discrete-time setup, when in the latter scheme each sensor transmits one-bit messages ($b^k = 2$) at every observation time ($r = 1$). Under these assumptions, it has been reported (see [14], [32]) that Q-CUSUM typically performs better than the min-CUSUM (see [32]) and that \mathcal{M} performs worse than Q-CUSUM in the case of large sensor networks.

In the case that each sensor observes a Brownian path, it was shown numerically in [18] that the performance of Q-CUSUM does not improve with a very low communication period r . Moreover, it was shown that min-CUSUM and Q-CUSUM have essentially the same performance when in the latter each sensor transmits one-bit messages ($b^k = 2$).

The above comparisons offer some important insights, however they do not take into account the fact that *the compared schemes have very different communication activities*. Therefore, these comparisons may not be very informative, since the goal in the decentralized setup is to optimize the performance of the detection rule while controlling the overall communication load in the network.

3.4. D-CUSUM; A novel scheme. From the previous discussion it is clear that it remains an open problem to find an asymptotically optimal and efficient decentralized detection rule, whose communication rate can be controlled and whose efficiency can be preserved even with a low communication rate or a large number of sensors. Our contribution in this work is that we propose and analyze a novel decentralized detection rule with these characteristics.

The proposed scheme is based on threshold quantization of the local log-likelihood ratios and a CUSUM-like rule at the fusion center, just like Q-CUSUM. Its difference is that the sensors communicate *asynchronously* with the fusion center, at *random* instead of deterministic times, in particular at two-sided exit times of the local log-likelihood ratios. Due to this characteristic, the proposed detection rule turns out to be asymptotically optimal, contrary to Q-CUSUM. Actually, it can attain a much stronger (second-order) asymptotic optimality than Mei's scheme, whereas its asymptotic optimality is preserved even with an asymptotically low communication rate and a large

sensor network. Finally, these theoretical properties are accompanied by a very good performance in practice, as our simulation experiments suggest.

The design and analysis of the proposed scheme differ significantly depending on whether the sensors take discrete or continuous time observations. Therefore, we will consider these two cases separately in the next two sections. However, we will see that the properties of the suggested rule, which we will call D-CUSUM, turn out to be very similar under both setups.

4. D-CUSUM in Continuous-Time. In this section, $t \in [0, \infty)$ and each log-likelihood ratio process $\{u_t^k\}_{t \geq 0}$ is assumed to have *continuous paths* for every $k = 1, \dots, K$. Moreover, we assume that condition (2.8) is satisfied, which guarantees the optimality of the centralized CUSUM test, \mathcal{S} . We also have the following closed-form expressions (see [17], [3])

$$(4.1) \quad \begin{aligned} \gamma &= \mathbb{E}_\infty[-u_{\mathcal{S}}] = \mathbb{E}_\infty[\langle u \rangle_{\mathcal{S}}] = e^\nu - \nu - 1 \\ \mathcal{J}[\mathcal{S}] &= \mathbb{E}_0[u_{\mathcal{S}}] = \mathbb{E}_0[\langle u \rangle_{\mathcal{S}}] = e^{-\nu} + \nu - 1. \end{aligned}$$

Moreover, we assume that the sensor processes are independent, thus (3.3) holds. However, we will see that this assumption can be removed in the case that the sensors observe correlated Brownian motions.

Our goal in this section is to describe the continuous-time version of the proposed scheme and compare its performance with that of the optimal centralized CUSUM test.

We suggest that each sensor k communicates with the fusion center at the sequence of $\{\mathcal{T}_t^k\}$ -stopping times that is defined by the following recursion:

$$(4.2) \quad \tau_n^k := \inf\{t > \tau_{n-1}^k : u_t^k - u_{\tau_{n-1}^k}^k \notin (-\underline{\Delta}^k, \bar{\Delta}^k)\}, \quad n \in \mathbb{N}; \quad \tau_0^k := 0,$$

where the thresholds $\bar{\Delta}^k, \underline{\Delta}^k$ are fixed, positive constants, known to sensor k and the fusion center. We denote by $\ell_n^k := u_{\tau_n^k}^k - u_{\tau_{n-1}^k}^k$ the accumulated log-likelihood ratio at sensor k in the time-interval $[\tau_{n-1}^k, \tau_n^k]$. Then, due to the path-continuity of $\{u_t^k\}$, it is clear that each ℓ_n^k will be exactly equal to either $\bar{\Delta}^k$ or $-\underline{\Delta}^k$. Therefore, if the fusion center receives at τ_n^k the following *one-bit* message by sensor k

$$(4.3) \quad z_n^k := \begin{cases} 1, & \text{if } \ell_n^k = \bar{\Delta}^k \\ -1, & \text{if } \ell_n^k = -\underline{\Delta}^k \end{cases}$$

it learns the *exact* value of ℓ_n^k , since $\ell_n^k = \bar{\Delta}^k \mathbb{1}_{\{z_n^k=1\}} - \underline{\Delta}^k \mathbb{1}_{\{z_n^k=-1\}}$. As a result, the fusion center is able to recover u_t^k at every communication time $t = \tau_n^k$, since $u_{\tau_n^k}^k = \ell_1^k + \dots + \ell_n^k$. Then, a natural approximation for u_t^k at

some arbitrary time t is the corresponding most recently reproduced value, i.e.

$$(4.4) \quad \tilde{u}_t^k := \sum_{n=1}^{m_t^k} \ell_n^k, \quad m_t^k := \max\{n : \tau_n^k \leq t\}.$$

Finally, mimicking the centralized CUSUM rule, we propose the following detection rule at the fusion center

$$(4.5) \quad \tilde{\mathcal{S}} := \inf\{t \geq 0 : \tilde{y}_t \geq \tilde{\nu}\}, \text{ where } \tilde{y}_t := \tilde{u}_t - \inf_{0 \leq s \leq t} \tilde{u}_s, \quad \tilde{u}_t := \sum_{k=1}^K \tilde{u}_t^k$$

and the threshold $\tilde{\nu}$ is chosen so that $\mathbb{E}_\infty[-u_{\tilde{\mathcal{S}}}] = \gamma$.

4.1. Design and implementation. The proposed scheme has a number of practical advantages. First of all, the fusion statistic $\{\tilde{y}_t\}$ is piecewise-constant and needs to be updated only at the communication times from the sensors, according to the following convenient recursion formula:

$$(4.6) \quad \tilde{y}_{\tau_n^k} = (\tilde{y}_{\tau_n^k-})^+ + \bar{\Delta}^k \mathbb{1}_{\{z_n^k=1\}} - \underline{\Delta}^k \mathbb{1}_{\{z_n^k=-1\}}.$$

In other words, whenever it receives a message from sensor k , the fusion center simply needs to add $\bar{\Delta}^k$ or $-\underline{\Delta}^k$ to the positive part of the current value of its statistic, $\{\tilde{y}_t\}$. Compare this with the centralized, continuous-time CUSUM statistic $\{y_t\}$, which does not have this nice property (unless the sensors observe Brownian motions) and whose calculation at the fusion center requires high-frequency transmission of “infinite-bit” messages from the sensors.

The thresholds $\bar{\Delta}^k, \underline{\Delta}^k$ control the communication rate of sensor k , thus they should ideally be chosen in order to attain target values for the expected inter-communication times, $\mathbb{E}_0[\tau_n^k - \tau_{n-1}^k]$ and $\mathbb{E}_\infty[\tau_n^k - \tau_{n-1}^k]$. However, since these expectations in general depend on n , we propose instead to select $\bar{\Delta}^k, \underline{\Delta}^k$ to attain target values for $\mathbb{E}_0[\ell_n^k]$ and $\mathbb{E}_\infty[-\ell_n^k]$.

These quantities represent the expected accumulated Kullback-Leibler divergences between the post and pre-change measure in the path of ξ^k during the time-interval $[\tau_{n-1}^k, \tau_n^k]$ and they do not depend on n , since $\mathbb{E}_0[\ell_n^k] = s(\bar{\Delta}^k, \underline{\Delta}^k)$ and $-\mathbb{E}_\infty[\ell_n^k] = s(\underline{\Delta}^k, \bar{\Delta}^k)$, where

$$(4.7) \quad s(x, y) = \frac{x(e^y - 1) - y(e^{x+y} - e^y)}{e^{x+y} - 1}.$$

In this way, the specification of $\bar{\Delta}^k, \underline{\Delta}^k$ requires only the solution of a system of two non-linear equations.

4.2. *Asymptotic optimality.* From the previous discussion it should be clear that, from a practical point of view, D-CUSUM is much more preferable than the corresponding centralized CUSUM. Our goal in this section is to show that it also has excellent performance characteristics, making any additional benefit of the optimal centralized CUSUM test negligible relative to its implementation cost. The following theorem is crucial in this direction, as it provides a very useful, non-asymptotic upper bound on the performance loss of the proposed detection structure.

THEOREM 1. *For any γ and $\{\bar{\Delta}^k, \underline{\Delta}^k\}_{1 \leq k \leq K}$ we have*

$$(4.8) \quad \mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] \leq 4K \Delta_{\max},$$

where $\Delta_{\max} := \max_{1 \leq k \leq K} \{\bar{\Delta}^k, \underline{\Delta}^k\}$.

PROOF. The proof is presented in Appendix A. □

Let us discuss the implications of Theorem 1. The bound provided in (4.8) implies that for any fixed thresholds $\{\bar{\Delta}^k, \underline{\Delta}^k\}$ and any number of sensors K , the performance loss of $\tilde{\mathcal{S}}$ is bounded as $\gamma \rightarrow \infty$, in other words $\tilde{\mathcal{S}}$ is asymptotically optimal of *order-2*. There are also interesting conclusions that we can draw for the case of a large sensor-network ($K \rightarrow \infty$). In particular, if we let $K \rightarrow \infty$ and $\Delta_{\max} \rightarrow 0$ so that $K\Delta_{\max} = \mathcal{O}(1)$, then $\tilde{\mathcal{S}}$ remains asymptotically optimal of order-2. In other words, with a very large sensor network, second-order optimality is preserved if there is a sufficiently high rate of communication from the sensors to the fusion center. Of course, we can do even better by letting $K \rightarrow \infty$ and $\Delta_{\max} \rightarrow 0$ so that $K\Delta_{\max} = o(1)$, in which case D-CUSUM becomes asymptotically optimal of order-3 and its distance from the optimal centralized CUSUM vanishes asymptotically. However, since we want to avoid the frequent communication activity that is induced by letting $\Delta_{\max} \rightarrow 0$, it is more interesting to see that $\tilde{\mathcal{S}}$ remains asymptotically optimal (of order-1) in an asymptotically large sensor network ($K \rightarrow \infty$) *and/or* under an asymptotically low communication rate ($\Delta_{\max} \rightarrow \infty$), as long as $K\Delta_{\max} = o(\log \gamma)$. Indeed, from (4.1) and (4.8) we have

$$(4.9) \quad \frac{\mathcal{J}[\tilde{\mathcal{S}}]}{\mathcal{J}[\mathcal{S}]} = 1 + \frac{\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}]}{\mathcal{J}[\mathcal{S}]} \leq 1 + \frac{4K\Delta_{\max}}{e^{-\nu} + \nu - 1}$$

and our claim now follows from (4.1), which implies that $\nu = \log \gamma + o(1)$.

4.3. *The case of correlated sensors.* Going over the proof of Theorem 1 in Appendix A, we realize that the assumption of independence across sensors is needed only to the extent that it guarantees a decomposition of the form $u_t = \sum_{k=1}^K u_t^k$, where $\{u_t^k\}$ is an \mathcal{F}_t^k -adapted process with continuous paths. Indeed, we did not use at all the fact that $\{u_t^k\}$ is the local log-likelihood ratio at sensor k . This implies that the previous results (and the corresponding analysis) are valid even for sensors with correlated dynamics, as long as such a decomposition is possible. This is for example the case when the sensors observe correlated Brownian motions before and after the change, i.e.

$$(4.10) \quad \xi_t^k = \sum_{j=1}^K \sigma_{kj} W_t^j + \mathbb{1}_{\{t > \tau\}} \mu^k t, \quad t \geq 0, \quad k = 1, \dots, K,$$

where (W^1, \dots, W^K) is a standard K -dimensional Wiener process, $\mu = [\mu^1, \dots, \mu^K]'$ a K -dimensional real vector and $\sigma := [\sigma_{ij}]$ a square matrix of dimension K so that the diffusion coefficient matrix $\Sigma = \sigma\sigma'$ is invertible. Then, we can write $u_t = \sum_{k=1}^K [b^k \xi_t^k - 0.5 \mu^k b^k t]$, where $b = [b^1, \dots, b^K]'$ is $\Sigma^{-1}\mu$, and Theorem 1 remains valid as long as we replace in the definition (4.2) of the stopping times (τ_n^k) the local log-likelihood ratio, $u_t^k = \mu^k \xi_t^k - 0.5 (\mu^k)^2 t$, by $b^k \xi_t^k - 0.5 \mu^k b^k t$.

5. D-CUSUM in Discrete-Time. In this section, $t \in \mathbb{N}$ and the increments $\{u_t^k - u_{t-1}^k\}_{t \in \mathbb{N}}$ are independent and identically distributed random variables with a finite second moment under both \mathbf{P}_0^k and \mathbf{P}_∞^k for every $1 \leq k \leq K$. Moreover, we assume that the sensor processes are independent, thus (3.3) is valid.

Our goal in this section is to describe the design and optimality properties of the proposed scheme in this discrete-time setup, emphasizing on the differences and similarities with the continuous-time setup.

5.1. *Proposed decentralized detection rule.* As in the previous section, we assume that sensor k communicates with the fusion center at the following stopping times

$$(5.1) \quad \tau_n^k := \inf\{t > \tau_{n-1}^k : u_t^k - u_{\tau_{n-1}^k}^k \notin (-\underline{\Delta}^k, \bar{\Delta}^k)\}, \quad n \in \mathbb{N}; \quad \tau_0^k := 0,$$

where $\bar{\Delta}^k, \underline{\Delta}^k$ are fixed, positive thresholds, and we set

$$\ell_n^k := u_{\tau_n^k}^k - u_{\tau_{n-1}^k}^k, \quad n \in \mathbb{N}.$$

However, unlike the previous section, ℓ_n^k is not restricted to the binary set $\{\bar{\Delta}^k, -\underline{\Delta}^k\}$, but takes values in $(-\infty, -\underline{\Delta}^k] \cup [\bar{\Delta}^k, \infty)$. As a result, the fusion

center cannot recover ℓ_n^k exactly when sensor k can transmit only a finite number of bits. This additional source of performance degradation obligates us to modify the scheme we proposed in continuous time, both in terms of the messages that are transmitted to the fusion center, as well as in how the fusion center utilizes these messages.

First, we need to generalize the definition of transmitted messages in order to allow for the incorporation of richer alphabets. Thus, we assume that the alphabet at sensor k is of the form $\{-d^k, \dots, -1, 1, \dots, d^k\}$, where $d^k \geq 1$ is some positive integer, and we suggest that sensor k transmits at time τ_n^k the following message to the fusion center

$$(5.2) \quad z_n^k = \begin{cases} j, & \text{if } \bar{\epsilon}_{j-1}^k \leq \ell_n^k - \bar{\Delta}^k < \bar{\epsilon}_j^k \\ -j, & \text{if } -\underline{\epsilon}_j^k < \ell_n^k + \underline{\Delta}^k \leq -\underline{\epsilon}_{j-1}^k \end{cases}$$

where $j = 1, \dots, d^k$. The thresholds $\{\bar{\epsilon}_j^k, \underline{\epsilon}_j^k, j = 1, \dots, d^k - 1\}$ are selected so that

$$(5.3) \quad \begin{aligned} P_0(\ell_n^k - \bar{\Delta}^k > \bar{\epsilon}_j^k \mid \ell_n^k > \bar{\Delta}^k) &= 1 - \frac{j}{d^k} \\ P_\infty(\ell_n^k + \underline{\Delta}^k < -\underline{\epsilon}_j^k \mid \ell_n^k < -\underline{\Delta}^k) &= 1 - \frac{j}{d^k}, \end{aligned}$$

whereas $\bar{\epsilon}_0^k = \underline{\epsilon}_0^k := 0$, $\bar{\epsilon}_{d^k}^k := \text{essup } u_1^k$ and $\underline{\epsilon}_{d^k}^k := \text{essup } (-u_1^k)$. In this way, the overshoot $\ell_n^k - \bar{\Delta}^k$ is equally likely to lie in each interval $(\bar{\epsilon}_{j-1}^k, \bar{\epsilon}_j^k)$ and, similarly, $-(\ell_n^k + \underline{\Delta}^k)$ is equally likely to lie in each interval $(\underline{\epsilon}_{j-1}^k, \underline{\epsilon}_j^k)$, where $j = 1, \dots, d^k$.

The second modification is in the way the fusion center approximates the log-likelihood ratio u_t^k at some arbitrary time t . In the previous section, we used the value of u^k at the most recent communication time from sensor k , recall (4.4). This approximation was possible due to the path-continuity of $\{u_t^k\}$, which allowed the fusion center to learn the *exact* value of each binary random variable ℓ_n^k . Since this is no longer the case in discrete time, we now define

$$(5.4) \quad \tilde{u}_t^k := \sum_{n=1}^{m_t^k} \tilde{\ell}_n^k, \quad m_t^k := \max\{n : \tau_n^k \leq t\},$$

where $\tilde{\ell}_n^k$ is now an *approximation* of ℓ_n^k that relies on $\tau_n^k - \tau_{n-1}^k$ and z_n^k , the information that becomes available to the fusion center at time τ_n^k . A straightforward choice is to define $\tilde{\ell}_n^k$ as the log-likelihood ratio of the pair $(\tau_n^k - \tau_{n-1}^k, z_n^k)$. Unfortunately, this is not possible in practice, since the

distribution of the inter-communication times $(\tau_n^k - \tau_{n-1}^k)_{n \in \mathbb{N}}$ is typically intractable. For this reason, we follow a *partial likelihood* approach and define $\tilde{\ell}_n^k$ as the log-likelihood ratio of z_n^k :

$$(5.5) \quad \tilde{\ell}_n^k := \sum_{j=1}^{d^k} \left[\bar{\Lambda}_j^k \mathbb{1}_{\{z_n^k=j\}} - \underline{\Lambda}_j^k \mathbb{1}_{\{z_n^k=-j\}} \right], \quad \text{where}$$

$$(5.6) \quad \bar{\Lambda}_j^k := \log \frac{P_0(z_n^k = j)}{P_\infty(z_n^k = j)}, \quad -\underline{\Lambda}_j^k := \log \frac{P_0(z_n^k = -j)}{P_\infty(z_n^k = -j)}.$$

Note that since $\{u_t^k\}_{t \in \mathbb{N}}$ is a random walk, the messages $(z_n^k)_{n \in \mathbb{N}}$ are iid, which is why $\bar{\Lambda}_j^k$ and $\underline{\Lambda}_j^k$ do not depend on n .

The proposed detection rule at the fusion center will then have the same form as in (4.5), i.e. $\tilde{S} := \inf\{t \in \mathbb{N} : \tilde{y}_t \geq \tilde{\nu}\}$, where $\tilde{y}_t := \tilde{u}_t - \inf_{0 \leq s \leq t} \tilde{u}_s$, $\tilde{u}_t := \sum_{k=1}^K \tilde{u}_t^k$, and the threshold $\tilde{\nu}$ is assumed to satisfy $\mathbb{E}_\infty[-u_{\tilde{S}}] = \gamma$.

5.2. Design and implementation. As in the previous section, $\bar{\Delta}^k$ and $\underline{\Delta}^k$ control the communication activity of sensor k . Since each $\{u_t^k\}$ is a random walk, we have $\mathbb{E}_0[\tau_n^k - \tau_{n-1}^k] = \mathbb{E}_0[\tau_1^k]$ and $\mathbb{E}_\infty[\tau_n^k - \tau_{n-1}^k] = \mathbb{E}_\infty[\tau_1^k]$, therefore $\bar{\Delta}^k$ and $\underline{\Delta}^k$ can be selected so that target values for $\mathbb{E}_0[\tau_1^k]$ and $\mathbb{E}_\infty[\tau_1^k]$ be attained. This can be done for example with a stochastic approximation algorithm, since the latter quantities are not in general available in closed-form but can be approximated with simulations.

According to (5.3), thresholds $\{\epsilon_j^k\}$ and $\{\epsilon_{-j}^k\}$ are percentiles of $\ell_1^k - \bar{\Delta}^k$ and $-(\ell_1^k + \underline{\Delta}^k)$, thus their computation simply requires the simulation of ℓ_1^k .

After having specified $\bar{\Delta}^k, \underline{\Delta}^k$ and $\{\epsilon_j^k, \epsilon_{-j}^k, j = 1, \dots, d^k - 1\}$, the next step is to compute $\{\bar{\Lambda}_j^k, \underline{\Lambda}_j^k\}$, which can also be done with simulations, since these quantities do not admit closed-form expressions. However, this is not a straightforward task if one uses their definition in (5.6), since this expression requires the simulation of rare events. Nevertheless, we can overcome this problem using the following lemma.

LEMMA 1. *For every $1 \leq j \leq d^k$,*

$$(5.7) \quad \begin{aligned} \bar{R}_j^k &:= \bar{\Lambda}_j^k - \bar{\Delta}_j^k = -\log \mathbb{E}_0[e^{-(\ell_n^k - \bar{\Delta}_j^k)} | z_n^k = j], \\ \underline{R}_j^k &:= \underline{\Lambda}_j^k - \underline{\Delta}_j^k = -\log \mathbb{E}_\infty[e^{\ell_n^k + \underline{\Delta}_j^k} | z_n^k = -j]. \end{aligned}$$

where

$$(5.8) \quad \bar{\Delta}_j^k := \bar{\Delta}^k + \epsilon_{j-1}^k, \quad \underline{\Delta}_j^k := \underline{\Delta}^k + \epsilon_{-j-1}^k.$$

PROOF. The proof of this lemma can be found in Appendix B. \square

5.3. *The error due to quantization.* The transmission of z_n^k requires the communication of $\lceil \log_2(2d^k) \rceil = 1 + \lceil \log_2 d^k \rceil$ bits. When $d^k = 1$, z_n^k is a one-bit message that simply informs the fusion center whether ℓ_n^k is above $\bar{\Delta}^k$ or below $-\underline{\Delta}^k$. When $d^k \geq 2$, z_n^k also informs the fusion center regarding the size of the overshoots $\ell_n^k - \bar{\Delta}^k$ or $\ell_n^k + \underline{\Delta}^k$, mitigating in this way the error due to quantization. The following lemma makes this observation more precise, providing an upper bound for the quantization error in an arbitrary transmission to the fusion center.

LEMMA 2. *For any k, n we have*

$$(5.9) \quad \ell_n^k - \tilde{\ell}_n^k \leq \eta_n^k := \sum_{j=1}^{d^k} \left[(\ell_n^k - \bar{\Delta}_j^k) \mathbb{1}_{\{z_n^k=j\}} + \underline{R}_j^k \mathbb{1}_{\{z_n^k=-j\}} \right].$$

and $\mathbb{E}_0[\eta_n^k] \leq \theta^k$, where θ^k is a constant that depends on d^k so that $\theta^k \rightarrow 0$ as $d^k \rightarrow \infty$.

PROOF. The proof of this lemma can be found in Appendix B. □

Although it may seem counterintuitive at first, we expect that when $d^k = 1$, a very high rate of communication for sensor k (small $\{\bar{\Delta}^k, \underline{\Delta}^k\}$) will not be desirable, since it may lead to a fast accumulation of quantization error. However, as the previous lemma suggests, this will not be the case with larger alphabets. This intuition will be supported by our asymptotic analysis in the next subsection.

Quantizing the overshoots presents another very important characteristic that should be mentioned. Clearly the statistical behavior of the overshoots depends on the two main thresholds $\bar{\Delta}^k, \underline{\Delta}^k$, which control the average period the sensor communicates with the fusion center. However, this dependency is only minor since the pdf of the overshoots converges to some limiting pdf as these thresholds become large. In other words, quantizing the overshoots is like quantizing a random variable with (almost) fixed statistics. Consequently, the mean square quantization error (or any other similar quality measure), for fixed number of bits, will be (almost) independent from the two thresholds $\bar{\Delta}^k, \underline{\Delta}^k$.

Let us now turn to the classical quantization scheme (3.4) employed by Q-CUSUM where quantization is applied on the value of $u_{nr}^k - u_{(n-1)r}^k$, where nr denotes the times the sensor communicates with the fusion center and r the corresponding period. It is very simple to realize that for fixed number of bits, if we increase the period r , the mean square quantization error will *increase* since the difference $u_{nr}^k - u_{(n-1)r}^k$ will involve a larger sum of

i.i.d. random variables. This becomes particularly obvious when these random variables are bounded, in which case the support of the sum increases linearly with r and we are asked, with the same number of bits, to quantize a larger range of values. This suggests that if we like to communicate with the fusion center at a smaller rate and preserve the same number of bits, this will inflict larger quantization errors and therefore additional performance degradation. This is clearly not the case with the quantization scheme we adopt for D-CUSUM since increasing $\bar{\Delta}^k, \underline{\Delta}^k$ (to reduce the communication rate) leaves the mean square quantization error almost intact.

5.4. Asymptotic Optimality. Our goal in this section is to draw conclusions regarding the optimal design and asymptotic behavior of $\tilde{\mathcal{S}}$ as $\gamma \rightarrow \infty$. For simplicity of presentation, we assume that there is a quantity Δ so that $\bar{\Delta}^k, \underline{\Delta}^k = \Theta(\Delta)$ as $\Delta, \bar{\Delta}^k, \underline{\Delta}^k \rightarrow \infty$ for all $1 \leq k \leq K$, which means that the rate of communication, before and after the change, for all sensors is of the same order. Furthermore, we use $\theta := \max_{1 \leq k \leq K} \theta^k$ as the parameter that controls the quantization error in any transmission to the fusion center, where we recall that θ^k appears in Lemma 2.

THEOREM 2. *As $\gamma \rightarrow \infty$ we have*

$$(5.10) \quad 0 \leq \mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] \leq \theta \frac{\log \gamma}{\Theta(\Delta)} + K \Theta(\Delta).$$

PROOF. The optimality of the CUSUM test implies that $\mathcal{J}[\tilde{\mathcal{S}}] \geq \mathcal{J}[\mathcal{S}]$, therefore it suffices to prove the second inequality in (5.10). For both the optimum CUSUM \mathcal{S} and D-CUSUM $\tilde{\mathcal{S}}$ we know that $\mathcal{J}[\mathcal{S}] = \mathbb{E}_0[u_{\mathcal{S}}]$ and $\mathcal{J}[\tilde{\mathcal{S}}] = \mathbb{E}_0[u_{\tilde{\mathcal{S}}}]$, therefore we can write

$$(5.11) \quad \mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] = \mathbb{E}_0[u_{\tilde{\mathcal{S}}}] - \mathbb{E}_0[u_{\mathcal{S}}] = \mathbb{E}_0[u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}] + \mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}] - \mathbb{E}_0[u_{\mathcal{S}}].$$

Thus, we need to provide suitable bounds for the three terms in the right hand side of (5.11). Indeed, from Lemma 6 we have $\mathbb{E}_0[u_{\mathcal{S}}] \geq \log \gamma - K\Theta(1)$, whereas from Lemma 7 we have $\mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}] \leq \log \gamma + K\Theta(\Delta)$. Moreover, from Lemmas 8 and 9 we obtain

$$\mathbb{E}_0[u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}] \leq K\Theta(\Delta) + \theta \frac{\log \gamma}{\Theta(\Delta)}.$$

Applying these inequalities in (5.11) we obtain the desired result. The above lemmas, as well as some additional auxiliary results, are stated and proved in Appendices C and D. \square

Let us now discuss the implications of this theorem. First of all, we observe that when K and Δ are fixed, the inflicted performance loss of the proposed scheme relative to the centralized CUSUM is bounded, i.e. $\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] = \mathcal{O}(1)$ as $\gamma \rightarrow \infty$, only if $\theta^k \rightarrow 0$ so that $\theta^k \log \gamma = \mathcal{O}(1)$ for every $1 \leq k \leq K$.

From Lemma 2 we know that $\theta^k \rightarrow 0$ as $d^k \rightarrow \infty$, however it would be more interesting to have an explicit divergence rate for d^k in terms of γ . It is easy to do so when each u_1^k has bounded support, so that $\bar{\epsilon}_{d^k}^k < \infty$ and $\underline{\epsilon}_{d^k}^k < \infty$. Indeed, in this special case it follows from (5.3) (see also (B.10)) that $\theta^k = \mathcal{O}(d^k)$, which means that the proposed scheme achieves second-order asymptotic optimality, when K and Δ are fixed, only if $d^k \rightarrow \infty$ so that $d^k = \mathcal{O}(\log \gamma)$ for every $1 \leq k \leq K$. In other words, the number of bits required for each transmission of sensor k (for fixed K and Δ) should be of the order $1 + \mathcal{O}(\log \log \gamma)$ for every $1 \leq k \leq K$.

When we have low communication rate in the sense that $\Delta \rightarrow \infty$, second-order asymptotic optimality cannot be preserved even when $\theta \rightarrow 0$. However, it can be preserved in a large sensor network ($K \rightarrow \infty$), as long as there is a sufficiently *high* communication rate, that is $\Delta \rightarrow 0$ so that $K\Delta = \mathcal{O}(1)$, and additionally $\theta \rightarrow 0$ so that $\theta \log \gamma = \mathcal{O}(\Delta)$.

First-order asymptotic optimality is achieved under significantly less strict conditions. Indeed, from Theorem 2 and Lemma 6 we have

$$\frac{\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}]}{\mathcal{J}[\mathcal{S}]} \leq \frac{\theta}{\Theta(\Delta)} + \frac{\Theta(\Delta)}{\frac{\log \gamma}{K} + 1}.$$

Therefore, D-CUSUM is first-order asymptotically optimal when

$$(5.12) \quad \frac{\theta}{\Delta} \rightarrow 0; \quad \frac{\log \gamma}{K} \rightarrow \infty; \quad \text{and} \quad \Delta = o\left(\frac{\log \gamma}{K}\right).$$

If θ is bounded away from 0, i.e. if $\{d^k, 1 \leq k \leq K\}$ are fixed, then first-order asymptotic optimality of D-CUSUM *requires* a low communication rate, i.e. $\Delta \rightarrow \infty$, but at an order which is smaller than the order of the ratio $\log \gamma / K$ which must tend to ∞ . On the other hand, if $d^k \rightarrow \infty$ for every $1 \leq k \leq K$ so that $\theta \rightarrow 0$, first-order asymptotic optimality is possible even when Δ is fixed.

The best possible upper bound for the inflicted performance loss of D-CUSUM is

$$(5.13) \quad \mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] \leq \mathcal{O}(\sqrt{K\theta \log \gamma})$$

and it is attained when Δ, θ and K are selected to equate, in order of mag-

nitude, the two terms of the upper bound in (5.10). This happens when

$$(5.14) \quad \Delta = \Theta \left(\sqrt{\frac{\theta \log \gamma}{K}} \right).$$

Therefore, when K and θ are fixed and Δ is selected using the previous relationship, D-CUSUM has the same asymptotic performance loss as Mei's scheme, that is $\mathcal{J}[\tilde{\mathcal{S}}] - \mathcal{J}[\mathcal{S}] = \mathcal{O}(\sqrt{\log \gamma})$. However, its performance is significantly improved when $\theta \rightarrow 0$, even if $K, \Delta \rightarrow \infty$.

We conclude that D-CUSUM in discrete time has essentially the same asymptotic behavior as in continuous time, as long as the parameter that controls the overall quantization error, θ , goes to zero as $\gamma \rightarrow \infty$. However, the very interesting implication of our analysis is that this can be achieved with a very small number of bits transmitted at each communication by each sensor. This is also verified by a small simulation experiment that we present in the next section.

Finally, we should note that when each sensor k transmits to the fusion center (with an “infinite-bit” message) the exact value of ℓ_n^k at each time τ_n^k , then (5.10) remains valid with $\theta = 0$, similarly to the continuous-time setup of the previous section. Of course, the resulting detection rule will be more efficient than any version of D-CUSUM that uses finite-bit messages.

5.5. Simulation experiments and comparisons. In this section, we illustrate the performance characteristics of $\tilde{\mathcal{S}}$ in the case of Gaussian observations. Thus, we assume that each sensor k takes independent, normally distributed observations with variance 1 and mean that changes from 0 to $\mu_k \neq 0$ at time τ . Therefore, the local log-likelihood ratio process in this example is a Gaussian random walk, in particular

$$(5.15) \quad u_t^k = \sum_{n=1}^t \left[\mu_k \xi_n^k - \frac{\mu_k^2}{2} \right], \quad t \in \mathbb{N}.$$

Note that the distribution of u_1^k under \mathbf{P}_0 is the same as that of $-u_1^k$ under \mathbf{P}_∞ , consequently the pair (τ_1^k, ℓ_1^k) will have the same distribution under \mathbf{P}_0 and \mathbf{P}_∞ .

We set $\bar{\Delta}^k = \underline{\Delta}^k = \Delta^k$ and $\bar{\epsilon}_j^k = \underline{\epsilon}_j^k = \epsilon_j^k$ for every $j = 1, \dots, d^k - 1$, consequently it is also going to be $\bar{\Lambda}_j^k = \underline{\Lambda}_j^k = \Lambda_j^k$ for every $j = 1, \dots, d^k - 1$. Moreover, we assume that $\mu_k = \mu$, $d^k = d$ and that Δ^k is chosen so that $\mathbf{E}_0[\tau_1^k] = r$ for every $1 \leq k \leq K$.

Our goal is to compare D-CUSUM $\tilde{\mathcal{S}}$ with Q-CUSUM $\hat{\mathcal{S}}$, which was defined in (3.5), when both rules use the same resources, i.e. the same number of bits per communication (one or two) and the same (average) rate

TABLE 1
Thresholds and Log-Likelihood Ratios

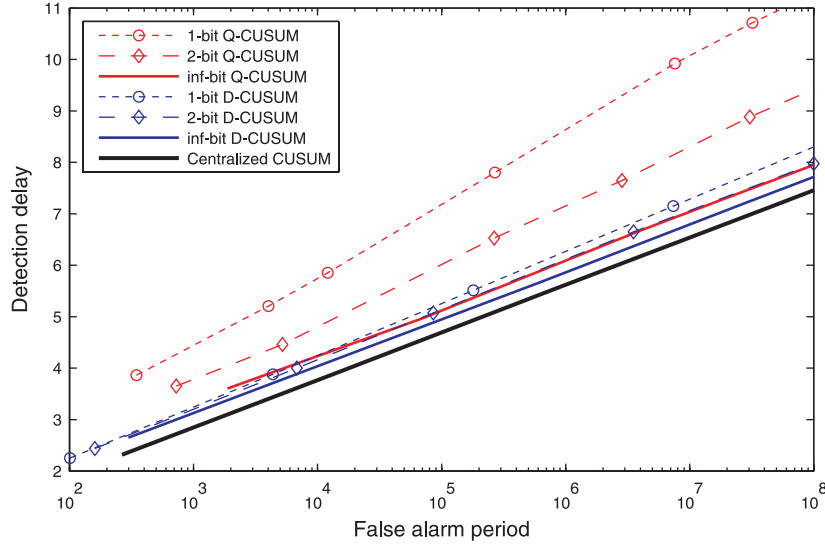
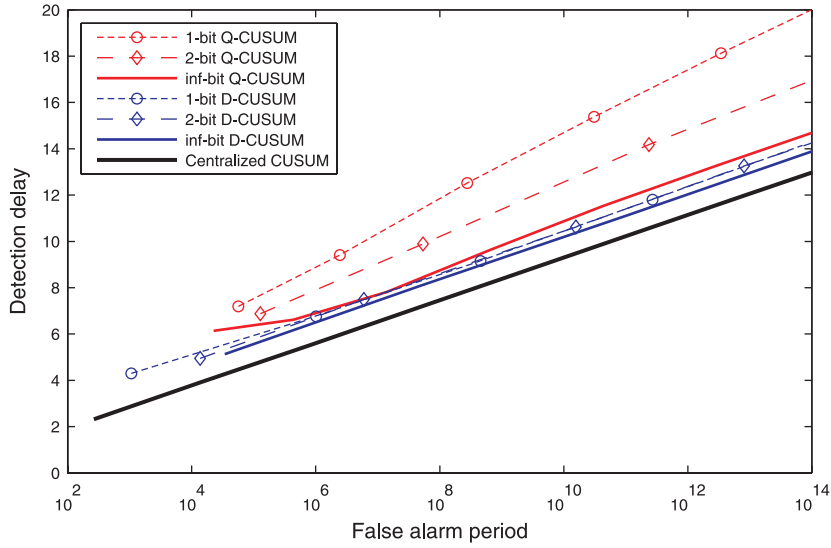
	Δ_1^k	Λ_1^k		Δ_1^k	Δ_2^k	Λ_1^k	Λ_2^k
$r = 3, \mu = 1$	1.287	1.87	$r = 3, \mu = 1$	1.287	1.87	1.54	2.94
$r = 6, \mu = 1$	2.54	3.12	$r = 6, \mu = 1$	2.54	3.12	2.80	3.62

(a) $d^k = 1$ (b) $d^k = 2$

of communication. Notice that such a legitimate comparison is not possible with decentralized rules, such as Mei's scheme, that cannot control explicitly their transmission rate. Of course, the ultimate benchmark is the centralized CUSUM test, which requires transmission of the complete sensor observation at every sampling time.

For $\mu = 1$, we present in Table 1 the parameters required for the implementation of the D-CUSUM when the number of transmitted bits is $d = 1$ or $d = 2$ and the communication period is $r = 3$ or $r = 6$. Fig. 1 and Fig. 2 depict our simulation results. First of all, we observe that in all cases the operating characteristics of D-CUSUM $\tilde{\mathcal{S}}$ are essentially parallel with those of the optimal centralized CUSUM, \mathcal{S} . This is exactly the *order-2* asymptotic optimality that we established theoretically. On the contrary, the performance loss of Q-CUSUM $\hat{\mathcal{S}}$ diverges in every case except when we transmit information with an infinite number of bits. This is expected since, as we argued before, this scheme is not even asymptotically optimal of order-1. Moreover, we observe that in all cases D-CUSUM is significantly more efficient than Q-CUSUM, since with one or two bits it is either very close or outperforms the *infinite-bit* Q-CUSUM.

Finally, it is worth mentioning that the performance difference between the one-bit and the infinite-bit D-CUSUM is only a minor percentage of the overall detection delay. This is particularly apparent when the average communication period is large, (compare Fig. 1, where average period is $r = 3$, with Fig. 2, where average period $r = 6$). Indeed, since larger communication periods imply larger values for the main thresholds $\bar{\Delta}^k, \underline{\Delta}^k$, the performance loss due to the unobserved overshoots is reduced. As a result, we do not experience significant gains by having the sensors transmit additional bits to the fusion center, which suggests that D-CUSUM enjoys in practice second-order asymptotic optimality even with one-bit transmissions. This is similar to the continuous-time case where, as we proved in Section 4, one-bit transmissions suffice for exact second-order asymptotic optimality. Summarizing, we can say that simulation experiments corroborate the conclusions of our asymptotic analysis.


 FIG 1. Case of $K = 5$ sensors with communication period $r = 3$.

 FIG 2. Case of $K = 5$ sensors with communication period $r = 6$.

6. Conclusions. In this work, we formulated the problem of decentralized sequential change detection under a setup which takes into account quantization and rate constraints. We argued that this formulation is more appropriate for applications that rely on sensor networks, where the goal is

to design an efficient detection rule while controlling the overall communication load. Moreover, we presented existing decentralized schemes under a unified framework that highlighted the communication activities they require.

We suggested a novel decentralized detection rule, according to which a sensor communicates with the fusion center at two-sided exit times of its local log-likelihood ratio process. The fusion center then uses a CUSUM-like rule in order to detect the change. The design and analysis of this scheme depend heavily on whether the sensors observe continuous-time processes with continuous-paths or random walks in discrete time. However, in both cases we showed that, with an appropriate design, the proposed scheme has essentially the same behavior. More specifically, its performance loss with respect to the centralized CUSUM is bounded (*second order* asymptotic optimality) for any fixed communication rate and any number of sensors, as long the quantization error associated with each transmission is small. On the other hand, first-order asymptotic optimality is preserved even with a low communication rate and a large number of sensors.

Additionally, we illustrated with simulation experiments that the proposed scheme is very efficient in practice and that it performs significantly better than the corresponding decentralized scheme that communicates messages of the same information content, at the same rate, but at *deterministic* times.

Finally, we should emphasize that we were able to remove the assumption of independence across sensors in the special case that the sensor processes are *correlated* Brownian motions. However, it remains an open problem to establish general, asymptotically optimal, decentralized detection rules when the sensors take correlated observations.

APPENDIX A

In what follows, we will use the following notation

$$(A.1) \quad \mathcal{S}_x = \inf\{t \geq 0 : y_t \geq x\}, \quad \tilde{\mathcal{S}}_x = \inf\{t \geq 0 : \tilde{y}_t \geq x\},$$

where $x > 0$. We also define the function

$$(A.2) \quad \psi(x) := \mathbb{E}_\infty[-u_{\mathcal{S}_x}] = \mathbb{E}_\infty[\langle u \rangle_{\mathcal{S}_x}], \quad x \geq 0,$$

and from (4.1) we conclude that

$$(A.3) \quad \psi(x) = e^x - x - 1, \quad x \geq 0.$$

Finally, as we have done so far, for any given γ we denote by ν and $\tilde{\nu}$ the thresholds for which $\mathbb{E}_\infty[-u_{\mathcal{S}_\nu}] = \mathbb{E}_\infty[-u_{\tilde{\mathcal{S}}_{\tilde{\nu}}}] = \gamma$. Before presenting the proof of Theorem 1 we need the following auxiliary lemma.

LEMMA 3. *For any $\gamma > 0$ we have the following double inequality which is true with probability 1*

$$(A.4) \quad \mathcal{S}_{\tilde{\nu}-2C} \leq \tilde{\mathcal{S}}_{\tilde{\nu}} \leq \mathcal{S}_{\tilde{\nu}+2C},$$

where $C := K\Delta_{\max}$ and $\Delta_{\max} := \max_{1 \leq k \leq K} \max\{\bar{\Delta}^k, \underline{\Delta}^k\}$.

PROOF. By definition, \tilde{u}^k coincides with u^k at the communication times $(\tau_n^k)_{n \in \mathbb{N}}$. Moreover, by (4.2) it is clear that the distance $|u^k - \tilde{u}^k|$ cannot be larger than $\max\{\bar{\Delta}^k, \underline{\Delta}^k\}$ between any two consecutive communication times. Thus, for any $t > 0$ we have

$$|u_t - \tilde{u}_t| \leq \sum_{k=1}^K |u_t^k - \tilde{u}_t^k| \leq K\Delta_{\max} = C.$$

As a result, for each $0 \leq s \leq t$ we have $u_t - C \leq \tilde{u}_t \leq u_t + C$,

$$\inf_{0 \leq s \leq t} u_s - C \leq \inf_{0 \leq s \leq t} \tilde{u}_s \leq \inf_{0 \leq s \leq t} u_s + C,$$

and applying these two inequalities in the definition of the statistic \tilde{y}_t in (4.5) we obtain $y_t - 2C \leq \tilde{y}_t \leq y_t + 2C$. Finally, due to this *pathwise* inequality, we immediately deduce that

$$\inf\{t : y_t - 2C \geq \tilde{\nu}\} \geq \inf\{t : \tilde{y}_t \geq \tilde{\nu}\} \geq \inf\{t : y_t + 2C \geq \tilde{\nu}\},$$

which is the claim of the lemma. \square

PROOF OF THEOREM 1. From the monotonicity of the quadratic variation process $\langle u \rangle$ and the previous lemma we have

$$(A.5) \quad \mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{S}_{\tilde{\nu}-2C}}] \leq \mathbb{E}_{\infty}[\langle u \rangle_{\tilde{\mathcal{S}}_{\tilde{\nu}}}] \leq \mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{S}_{\tilde{\nu}+2C}}].$$

Since $\mathbb{E}_{\infty}[\langle u \rangle_{\tilde{\mathcal{S}}_{\tilde{\nu}}}] = \mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{S}_{\nu}}] = \gamma$, we obtain

$$\mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{S}_{\tilde{\nu}-2C}}] \leq \mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{S}_{\nu}}] \leq \mathbb{E}_{\infty}[\langle u \rangle_{\mathcal{S}_{\tilde{\nu}+2C}}]$$

and using (A.2) we can write $\psi(\tilde{\nu} - 2C) \leq \psi(\nu) \leq \psi(\tilde{\nu} + 2C)$. From (A.3) it is clear that $\psi(\cdot)$ is strictly increasing, which implies that $|\nu - \tilde{\nu}| \leq 2C$. Therefore,

$$\begin{aligned} \mathcal{J}[\tilde{\mathcal{S}}_{\tilde{\nu}}] - \mathcal{J}[\mathcal{S}_{\nu}] &\leq \mathcal{J}[\mathcal{S}_{\tilde{\nu}+2C}] - \mathcal{J}[\mathcal{S}_{\nu}] \\ &= (e^{-\tilde{\nu}-2C} - e^{-\nu}) + (\tilde{\nu} - \nu) + 2C \leq 4C, \end{aligned}$$

where the first inequality follows from the previous lemma, the equality is due to (4.1) and the second inequality is due to the fact that $|\nu - \tilde{\nu}| \leq 2C$. \square

APPENDIX B

PROOF OF LEMMA 1. With a change of measure $\mathbf{P}_\infty \mapsto \mathbf{P}_0$ we have

$$\begin{aligned} \mathbf{P}_\infty(z_n^k = j) &= e^{-\bar{\Delta}_j^k} \mathbf{E}_0 \left[e^{-(\ell_n^k - \bar{\Delta}_j^k)} \mathbb{1}_{\{z_n^k = j\}} \right] \\ &= \mathbf{P}_0(z_n^k = j) e^{-\bar{\Delta}_j^k} \mathbf{E}_0 \left[e^{-(\ell_n^k - \bar{\Delta}_j^k)} \mid z_n^k = j \right]. \end{aligned}$$

Thus, from the definition of $\bar{\Lambda}_j^k$ in (5.6) we have

$$e^{-\bar{\Lambda}_j^k} = \frac{\mathbf{P}_\infty(z_n^k = j)}{\mathbf{P}_0(z_n^k = j)} = e^{-\bar{\Delta}_j^k} \mathbf{E}_0[e^{-(\ell_n^k - \bar{\Delta}_j^k)} \mid z_n^k = j]$$

and taking logarithms we obtain the first equality in (5.7), that is

$$\bar{\Lambda}_j^k - \bar{\Delta}_j^k = -\log \mathbf{E}_0[e^{-(\ell_n^k - \bar{\Delta}_j^k)} \mid z_n^k = j].$$

The second equality can be shown in a similar way. \square

PROOF OF LEMMA 2. In order to prove (5.9), we recall (5.5)-(5.6) and we argue as follows:

$$\begin{aligned} \ell_n^k - \tilde{\ell}_n^k &= \sum_{j=1}^{d^k} \left[(\ell_n^k - \bar{\Delta}_j^k - \bar{R}_j^k) \mathbb{1}_{\{z_n^k = j\}} + (\ell_n^k + \underline{\Delta}_j^k + \underline{R}_j^k) \mathbb{1}_{\{z_n^k = -j\}} \right] \\ \text{(B.1)} \quad &\leq \sum_{j=1}^{d^k} \left[(\ell_n^k - \bar{\Delta}_j^k) \mathbb{1}_{\{z_n^k = j\}} + \underline{R}_j^k \mathbb{1}_{\{z_n^k = -j\}} \right] =: \eta_n^k \end{aligned}$$

where the inequality is correct because $\bar{R}_j^k > 0$, which follows from (5.7), and $\ell_n^k + \underline{\Delta}_j^k < 0$ on $\{z_n^k = -j\}$. It remains to show that $\mathbf{E}_0[\eta_n^k]$ is bounded by a quantity that goes to 0 as $d^k \rightarrow \infty$. In order to do so, we denote by $\bar{\mathbf{P}}_0$ the probability measure \mathbf{P}_0 conditional on the event $\{z_n^k > 0\} = \{\ell_n^k > \bar{\Delta}^k\}$ and by $\bar{\mathbf{E}}_0$ the corresponding expectation. Then, taking expectations in (B.1) we have

$$\begin{aligned} \mathbf{E}_0[\eta_n^k] &= \sum_{j=1}^{d^k} \left[\mathbf{E}_0[(\ell_n^k - \bar{\Delta}_j^k) \mathbb{1}_{\{z_n^k = j\}}] + \underline{R}_j^k \mathbf{P}_0(z_n^k = -j) \right] \\ \text{(B.2)} \quad &\leq \sum_{j=1}^{d^k} \left[\bar{\mathbf{E}}_0[(\ell_n^k - \bar{\Delta}_j^k) \mathbb{1}_{\{z_n^k = j\}}] + \underline{R}_j^k \mathbf{P}_0(z_n^k = -j) \right]. \end{aligned}$$

and it suffices to upper bound appropriately every term in (B.2).

First of all, for any $1 \leq j \leq d^k - 1$ we have $\ell_n^k - \bar{\Delta}_j^k \leq \bar{\epsilon}_j^k - \bar{\epsilon}_{j-1}^k$ on $\{z_n^k = j\}$. Therefore, taking expectations and recalling (5.3) we have

$$\bar{\mathbb{E}}_0[(\ell_n^k - \bar{\Delta}_j^k) \mathbb{1}_{\{z_n^k=j\}}] \leq (\bar{\epsilon}_j^k - \bar{\epsilon}_{j-1}^k) \bar{\mathbb{P}}_0(z_n^k = j) \leq \frac{\delta^k}{d^k},$$

and consequently

$$(B.3) \quad \sum_{j=1}^{d^k-1} \bar{\mathbb{E}}_0[(\ell_n^k - \bar{\Delta}_j^k) \mathbb{1}_{\{z_n^k=j\}}] \leq (d^k - 1) \frac{\delta^k}{d^k} \leq \delta^k,$$

where δ^k is a finite constant defined as follows:

$$(B.4) \quad \delta^k := \max_{1 \leq j \leq d^k-1} \max\{\bar{\epsilon}_j^k - \bar{\epsilon}_{j-1}^k, \underline{\epsilon}_j^k - \underline{\epsilon}_{j-1}^k\}.$$

Since $\bar{\Delta}_{d^k}^k = \bar{\Delta}^k + \bar{\epsilon}_{d^k-1}^k$ and $\bar{\epsilon}_{d^k}^k = \text{essup } u_1^k$ we have:

$$(B.5) \quad \bar{\mathbb{E}}_0[(\ell_n^k - \bar{\Delta}_{d^k}^k) \mathbb{1}_{\{z_n^k=d^k\}}] = \int_{\bar{\epsilon}_{d^k-1}^k}^{\bar{\epsilon}_{d^k}^k} \bar{\mathbb{P}}_0(\ell_n^k - \bar{\Delta}^k > x) dx.$$

From [12, Theorem 4, Eq. (13)] we obtain

$$\begin{aligned} \bar{\mathbb{P}}_0(\ell_n^k - \bar{\Delta}^k > x) &\leq \frac{1}{\mathbb{E}_0[u_1^k]} \left(\frac{\bar{\Delta}^k + D}{\bar{\Delta}^k + x} \right) \mathbb{E}_0[(2u_1^k - x) \mathbb{1}_{\{u_1^k \geq x\}}] \\ &\leq \frac{2}{\mathbb{E}_0[u_1^k]} \left(1 + \frac{D}{\bar{\Delta}^k} \right) \mathbb{E}_0[u_1^k \mathbb{1}_{\{u_1^k \geq x\}}], \end{aligned}$$

where $D = \mathbb{E}_0[(u_1^k)^+]^2 / \mathbb{E}_0[u_1^k] \leq \mathbb{E}_0[(u_1^k)^2] / \mathbb{E}_0[u_1^k] < \infty$. Now note that for $\bar{\Delta}^k$ larger than some limiting value which is bounded away from 0, we have $(2/\mathbb{E}_0[u_1^k])(1 + D/\bar{\Delta}^k) \leq \bar{c}^k$, for properly selected constant \bar{c}^k that does not depend on $\bar{\Delta}^k$ and d^k . Using this upper bound in (B.5) and applying Fubini's theorem we obtain

$$\begin{aligned} \bar{\mathbb{E}}_0[(\ell_n^k - \bar{\Delta}^k) \mathbb{1}_{\{z_n^k=d^k\}}] &\leq \bar{c}^k \int_{\bar{\epsilon}_{d^k-1}^k}^{\bar{\epsilon}_{d^k}^k} \mathbb{E}_0[u_1^k \mathbb{1}_{\{u_1^k \geq x\}}] dx \\ (B.6) \quad &= \bar{c}^k \mathbb{E}_0[u_1^k (u_1^k - \bar{\epsilon}_{d^k-1}^k)^+] \leq \bar{c}^k \mathbb{E}_0[(u_1^k)^2 \mathbb{1}_{\{u_1^k > \bar{\epsilon}_{d^k-1}^k\}}]. \end{aligned}$$

For any $1 \leq j \leq d^k$, with a change of measure $P_0 \mapsto P_\infty$ we obtain

$$(B.7) \quad \begin{aligned} P_0(z_1^k = -j) &= E_\infty[e^{\ell_1^k} \mathbb{1}_{\{z_1^k = -j\}}] \\ &\leq P_\infty(z_1^k = -j) \leq \bar{P}_\infty(z_1^k = -j) = \frac{1}{d^k}, \end{aligned}$$

since $\ell_1^k \leq -\underline{\Delta}^k < 0$ on $\{z_1^k < 0\}$. Therefore, for any $1 \leq j \leq d^k - 1$ we have

$$P_0(z_1^k = -j) \underline{R}_j^k \leq \frac{\delta^k}{d^k},$$

since from (5.7) it follows that $\underline{R}_j^k \leq \underline{\epsilon}_j^k - \underline{\epsilon}_{j-1}^k \leq \delta^k$ on $\{z_n^k = j\}$ for $1 \leq j \leq d^k - 1$, and consequently we obtain

$$(B.8) \quad \sum_{j=1}^{d^k-1} P_0(z_1^k = -j) \underline{R}_j^k \leq (d^k - 1) \frac{\delta^k}{d^k} \leq \delta^k.$$

Finally, from an application of the conditional Jensen inequality in the second equality in (5.7) we have

$$(B.9) \quad \begin{aligned} P_0(z_n^k = -d^k) \underline{R}_{d^k}^k &\leq P_0(z_n^k = -d^k) E_\infty[\ell_n^k + \underline{\Delta}_{d^k}^k \mid z_n^k = -d^k] \\ &= \frac{P_0(z_n^k = -d^k)}{P_\infty(z_n^k = -d^k)} E_\infty[(\ell_n^k + \bar{\Delta}_{d^k}^k) \mathbb{1}_{\{z_n^k = -d^k\}}] \\ &\leq E_\infty[(\ell_n^k + \bar{\Delta}_{d^k}^k) \mathbb{1}_{\{z_n^k = -d^k\}}] \\ &\leq \underline{c}^k E_\infty[(u_1^k)^2 \mathbb{1}_{\{u_1^k \leq \underline{\epsilon}_{d^k-1}^k\}}], \end{aligned}$$

where \underline{c}^k is a constant term as $\bar{\Delta}^k, \underline{\Delta}^k, d^k \rightarrow \infty$. The second inequality follows from (B.7), whereas the third inequality can be shown in a similar way as (B.6).

Therefore, if we apply (B.3), (B.6), (B.8) and (B.9) to (B.2), we obtain $E_0[\eta_n^k] \leq \theta^k$, where

$$(B.10) \quad \theta^k := 2\delta^k + 2 \max \left\{ \bar{c}^k E_0[(u_1^k)^2 \mathbb{1}_{\{u_1^k \geq \bar{\epsilon}_{d^k-1}^k\}}], \underline{c}^k E_\infty[(u_1^k)^2 \mathbb{1}_{\{u_1^k \leq -\underline{\epsilon}_{d^k-1}^k\}}] \right\},$$

In order to complete the proof we need to show that $\theta^k \rightarrow 0$ as $d^k \rightarrow \infty$. This follows directly from the finiteness of the second moment of u_1^k and the definition of the thresholds $\{\bar{\epsilon}_j^k, \underline{\epsilon}_j^k\}$ in (5.3), which implies that $\delta^k \rightarrow 0$ as $d^k \rightarrow \infty$. \square

APPENDIX C

Our goal in this Appendix is to prove Lemma 4, which connects the threshold $\tilde{\nu}$ to the false-alarm period, γ . In order to provide an elegant proof of this result, we need to adopt an alternative representation of the fusion center policy (that we will use only in this Appendix). Indeed, since the implementation of $\tilde{\mathcal{S}}$ requires only the knowledge of the transmitted messages at the fusion center, it is possible to describe the fusion rule without any reference to the communication times $\{\tau_n^k\}$. Thus, let z_n be the n th message that arrives at the fusion center and k_n the corresponding identity of the sensor which transmitted this message. Of course, since time is discrete, there is non-zero probability that the fusion center may receive messages from two or more sensors concurrently. In this case, we enumerate the simultaneous messages in an arbitrary order and we keep the same order for the labels.

We can then describe the flow of information at the fusion center by the filtration $\{\mathcal{C}_n\}_{n \in \mathbb{N}}$, where $\mathcal{C}_n = \sigma((z_1, k_1) \dots, (z_n, k_n))$. For any $n \in \mathbb{N}$ we set

$$(C.1) \quad \begin{aligned} \phi_n &:= \log \frac{P_0(k_1, \dots, k_n)}{P_\infty(k_1, \dots, k_n)} \\ v_n &:= \log \frac{P_0(z_1, \dots, z_n | k_1, \dots, k_n)}{P_\infty(z_1, \dots, z_n | k_1, \dots, k_n)}. \end{aligned}$$

and recalling the definition of the log-likelihood ratios $\bar{\Lambda}_j^k, \underline{\Lambda}_j^k$ in (5.6), we have

$$(C.2) \quad v_n = \sum_{m=1}^n \sum_{j=1}^{d^{k_m}} \left[\bar{\Lambda}_j^{k_m} \mathbb{1}_{\{z_m=j\}} - \underline{\Lambda}_j^{k_m} \mathbb{1}_{\{z_m=-j\}} \right].$$

Then, the *number of messages* which the fusion center has received until an alarm is raised by D-CUSUM is given by the following $\{\mathcal{C}_n\}$ -stopping time:

$$(C.3) \quad \tilde{\mathcal{N}} = \inf\{n \in \mathbb{N} : v_n - \min_{m=1, \dots, n} v_m \geq \tilde{\nu}\}.$$

The process $\{v_n\}$ and the stopping time $\tilde{\mathcal{N}}$ are closely related to $\{\tilde{u}_t\}$ and $\tilde{\mathcal{S}}$, respectively. Their main difference is that $\{\tilde{u}_t\}$ and $\tilde{\mathcal{S}}$ are expressed in terms of “physical time”, whereas $\{v_n\}$ and $\tilde{\mathcal{N}}$ in terms of number of messages transmitted to the fusion center. If we denote by τ_n the time-instant at which the n th message arrives at the fusion center, then we can explicitly specify the following connection between these quantities: $\tilde{u}_{\tau_n} = v_n$ and $\tilde{\mathcal{S}} = \tau_{\tilde{\mathcal{N}}}$. In other words $\tilde{\mathcal{N}}$ denotes the number of received messages at the fusion center until stopping at time $\tilde{\mathcal{S}}$.

After these definitions, we can now prove Lemma 4, which connects $\tilde{\nu}$ to γ through an inequality that will be important for the performance analysis of $\tilde{\mathcal{S}}$.

LEMMA 4. *For any $\gamma > 0$ and $\tilde{\nu}$ so that $\mathbb{E}_\infty[-u_{\tilde{\mathcal{S}}}] = \gamma$, we have*

$$(C.4) \quad \tilde{\nu} \leq \log \gamma - \log(\bar{I}_\infty),$$

where \bar{I}_∞ is defined in (3.2).

PROOF. We first observe that

$$(C.5) \quad \gamma = \mathbb{E}_\infty[-u_{\tilde{\mathcal{S}}}] = K \bar{I}_\infty \mathbb{E}_\infty[\tilde{\mathcal{S}}] \geq \bar{I}_\infty \mathbb{E}_\infty[\tilde{\mathcal{N}}].$$

The second equality follows from an application of Wald's identity, whereas the inequality from the fact that $\tilde{\mathcal{N}} \leq K \tilde{\mathcal{S}}$. Indeed, the maximum number of received messages until stopping at $\tilde{\mathcal{S}}$ is obtained when at every time instant we have all sensors transmitting a message to the fusion center and this yields $K \tilde{\mathcal{S}}$.

From (C.5) it is clear that it suffices to prove $\mathbb{E}_\infty[\tilde{\mathcal{N}}] \geq e^{\tilde{\nu}}$. In order to do so, let us define the sequence $\{n_j\}$ of epochs where the CUSUM process $v_n - \min_{0 \leq m \leq n} v_m$ either returns to zero (restarts) or exceeds $\tilde{\nu}$. This is the classical way to write the CUSUM stopping time as a sum of a random number of components. Specifically, let us define

$$(C.6) \quad \begin{aligned} n_j &:= \inf\{n > n_{j-1} : v_n - v_{n_{j-1}} \notin (0, \tilde{\nu})\} \\ \mathcal{R} &:= \inf\{j \in \mathbb{N} : v_{n_j} - v_{n_{j-1}} \geq \tilde{\nu}\}. \end{aligned}$$

Then we clearly have $\tilde{\mathcal{N}} = n_{\mathcal{R}}$. Since from one epoch to the next we count at least one additional message, we trivially conclude that $\mathcal{R} \leq \tilde{\mathcal{N}}$ and, therefore, $\mathbb{E}_\infty[\mathcal{R}] \leq \mathbb{E}_\infty[\tilde{\mathcal{N}}]$. We can now claim that it suffices to show that

$$(C.7) \quad \mathbb{P}_\infty(\mathcal{R} > j) \geq (1 - e^{-\tilde{\nu}})^j, \quad \forall j \in \mathbb{N}.$$

In order to justify this claim, observe first that $\mathbb{E}_\infty[\tilde{\mathcal{N}}] < \infty$, since $\tilde{\mathcal{N}}$ is a CUSUM stopping time. As a result, $\mathbb{E}_\infty[\mathcal{R}]$ is finite as well and consequently (C.7) implies that

$$\mathbb{E}_\infty[\tilde{\mathcal{N}}] \geq \mathbb{E}_\infty[\mathcal{R}] = \sum_{j=0}^{\infty} \mathbb{P}_\infty(\mathcal{R} > j) \geq \sum_{j=0}^{\infty} (1 - e^{-\tilde{\nu}})^j \geq e^{\tilde{\nu}}.$$

In order to prove (C.7), we start with the following observation:

$$(C.8) \quad \begin{aligned} \mathbb{P}_\infty(\mathcal{R} > j) &= \mathbb{P}_\infty(\mathcal{R} > j-1; v_{n_j} - v_{n_{j-1}} \leq 0) \\ &= \mathbb{P}_\infty(\mathcal{R} > j-1) - \mathbb{P}_\infty(\mathcal{R} > j-1; v_{n_j} - v_{n_{j-1}} \geq \tilde{\nu}). \end{aligned}$$

Let us now set $A := \{\mathcal{R} > j - 1, v_{n_j} - v_{n_{j-1}} \geq \tilde{\nu}\}$. Then, it is clear that $A \in \mathcal{C}_{n_j}$ and with a change of measure $P_\infty \mapsto P_0$ we obtain

$$(C.9) \quad P_\infty(A) = \int_A \mathcal{L}_{n_j}^{-1} dP_0,$$

where for every $n \in \mathbb{N}$ we define

$$(C.10) \quad \mathcal{L}_n := e^{\phi_n + v_n}, \quad n \in \mathbb{N}.$$

We now argue as follows

$$(C.11) \quad \begin{aligned} P_\infty(A) &= \int_A \mathcal{L}_{n_{j-1}}^{-1} e^{-(\phi_{n_j} - \phi_{n_{j-1}}) - (v_{n_j} - v_{n_{j-1}})} dP_0 \\ &\leq e^{-\tilde{\nu}} \int_A \mathcal{L}_{n_{j-1}}^{-1} e^{-(\phi_{n_j} - \phi_{n_{j-1}})} dP_0 \\ &\leq e^{-\tilde{\nu}} \int_{\mathcal{R} > j-1} \mathcal{L}_{n_{j-1}}^{-1} e^{-(\phi_{n_j} - \phi_{n_{j-1}})} dP_0 \\ &= e^{-\tilde{\nu}} \int_{\mathcal{R} > j-1} \mathcal{L}_{n_{j-1}}^{-1} E_0[e^{-(\phi_{n_j} - \phi_{n_{j-1}})} | \mathcal{C}_{n_{j-1}}] dP_0. \end{aligned}$$

The first inequality is due to the fact that $v_{n_j} - v_{n_{j-1}} \geq \tilde{\nu}$ on the event A . The second inequality holds because $A \subset \{\mathcal{R} > j - 1\}$, whereas the last equality follows from the law of iterated expectation and the fact that $\{\mathcal{R} > j - 1\} \in \mathcal{C}_{n_{j-1}}$ and $\mathcal{L}_{n_{j-1}}^{-1}$ is a $\mathcal{C}_{n_{j-1}}$ -measurable random variable. Suppose now that

$$(C.12) \quad E_0[e^{-(\phi_{n_j} - \phi_{n_{j-1}})} | \mathcal{C}_{n_{j-1}}] = 1,$$

(a claim that we will prove shortly). Then, it is clear with a change of measure $P_\infty \mapsto P_0$ that (C.11) reduces to

$$(C.13) \quad P_\infty(A) \leq e^{-\tilde{\nu}} \int_{\mathcal{R} > j-1} \mathcal{L}_{n_{j-1}}^{-1} dP_0 = e^{-\tilde{\nu}} P_\infty(\mathcal{R} > j - 1).$$

Substituting the outcome of (C.13) in (C.8) and applying it repeatedly yields

$$P_\infty(\mathcal{R} > j) \geq (1 - e^{-\tilde{\nu}}) P_\infty(\mathcal{R} > j - 1) \geq (1 - e^{-\tilde{\nu}})^j.$$

Thus, in order to complete the proof it remains to justify (C.12). Since $\{e^{-\phi_n}\}_{n \in \mathbb{N}}$ is a $\{\mathcal{C}_n\}$ -martingale under P_0 , as a likelihood-ratio, and n_{j-1}, n_j are $\{\mathcal{C}_n\}$ -adapted stopping times, it suffices to show that the optional sampling theorem can be applied. More specifically, since $n_{j-1} \leq n_j$, we need to show that

$$(C.14) \quad E_0[e^{-\phi_{n_j}}] < \infty \quad \text{and} \quad \lim_{m \rightarrow \infty} E_0[e^{-\phi_m} \mathbb{1}_{\{n_j > m\}}] = 0.$$

Indeed, since ϕ_{n_j} is a \mathcal{C}_{n_j} -measurable random variable, with a change of measure $P_0 \mapsto P_\infty$ we have

$$(C.15) \quad E_0[e^{-\phi_{n_j}}] = E_\infty[e^{-\phi_{n_j}} \mathcal{L}_{n_j}] = E_\infty[e^{v_{n_j}}] < \infty,$$

and the last term is finite, since v_{n_j} cannot exceed $\tilde{\nu}$ by more than $\sum_{k=1}^K \bar{\Lambda}_j^k$, which is a finite quantity.

With the same change of measure argument we can show that the second condition in (C.14) is satisfied. Indeed,

$$\begin{aligned} E_0[e^{-\phi_m} \mathbb{1}_{\{n_j > m\}}] &= E_\infty[e^{-\phi_m} \mathcal{L}_m \mathbb{1}_{\{n_j > m\}}] = E_\infty[e^{-\phi_m} e^{v_m + \phi_m} \mathbb{1}_{\{n_j > m\}}] \\ &= E_\infty[e^{v_m} \mathbb{1}_{\{n_j > m\}}] \leq e^{\tilde{\nu}} P_\infty(n_j > m), \end{aligned}$$

where the last inequality is due to the fact that $v_m \leq \tilde{\nu}$ on $\{n_j > m\}$. Since n_j is almost surely finite, we have $P_\infty(n_j > m) \rightarrow 0$ as $m \rightarrow \infty$, which shows that the second condition in (C.14) is also satisfied and this completes the proof. \square

APPENDIX D

Our goal in this Appendix is to state and prove Lemmas 6, 7, 8 and 9, which are used in the proof of Theorem 2. In order to do so, we will need Lemma 5, which provides a very useful for our purposes, asynchronous version of Wald's identity. We recall that m_t^k is the number of messages that have been transmitted by sensor k to the fusion center up to time t and we denote by m_t the number of messages that have been transmitted by all sensors up to time t , i.e.

$$(D.1) \quad m_t^k := \max\{n \in \mathbb{N} : \tau_n^k \leq t\}, \quad m_t := \sum_{k=1}^K m_t^k.$$

LEMMA 5. *Consider a generic sequence $\{\zeta_n^k\}$, where each ζ_n^k is an arbitrary (Borel) function of the triplet $(\tau_n^k - \tau_{n-1}^k, z_n^k, \ell_n^k)$. Thus, $\{\zeta_n^k\}$ is a sequence of independent and identically distributed random variables under both P_0 (and P_∞). If \mathcal{T} is a P_0 -integrable $\{\mathcal{F}_t\}$ -stopping time and $E_0[|\zeta_1^k|] < \infty$, then*

$$(D.2) \quad E_0 \left[\sum_{n=1}^{m_{\mathcal{T}}^k + 1} \zeta_n^k \right] = E_0[m_{\mathcal{T}}^k + 1] E_0[\zeta_1^k].$$

If moreover $\zeta_n^k \geq 0$, then

$$(D.3) \quad \mathbb{E}_0 \left[\sum_{n=1}^{m_{\mathcal{T}}^k} \zeta_n^k \right] \leq (\mathbb{E}_0[m_{\mathcal{T}}^k] + 1) \mathbb{E}_0[\zeta_1^k].$$

Finally, if $|\zeta_n^k| \leq M^k$, where M^k is some finite constant, then

$$(D.4) \quad \mathbb{E}_0 \left[\sum_{n=1}^{m_{\mathcal{T}}^k} \zeta_n^k \right] \geq \mathbb{E}_0[m_{\mathcal{T}}^k] \mathbb{E}_0[\zeta_1^k] - 2M^k.$$

PROOF. The proof can be found in [6]. \square

LEMMA 6. *The optimum performance can be lower bounded as follows*

$$(D.5) \quad \mathbb{E}_0[u_{\mathcal{S}}] \geq \log \gamma - \Theta(K)$$

PROOF. Let us first define for any $r \geq 0$ the stopping times

$$T_r^+ = \inf\{t > 0 : u_t \geq r\}, \quad T_r^- = \inf\{t > 0 : -u_t \geq r\}.$$

Due to the representation of the CUSUM stopping time as a repeated SPRT with thresholds 0 and ν , we have the following well-known formula (see for example Siegmund, [37, Page 25]) for its expectation under \mathbb{P}_0 and \mathbb{P}_{∞}

$$(D.6) \quad \mathbb{E}_i[u_{\mathcal{S}}] = \frac{\mathbb{E}_i[u_{\mathcal{T}}]}{\mathbb{P}_i(u_{\mathcal{T}} \geq \nu)}, \quad i = 0, \infty,$$

where $\mathcal{T} = \min\{T_0^-, T_{\nu}^+\}$ is the SPRT stopping time with boundaries 0 and ν . Using (D.6) for $i = 0$, we can now write

$$(D.7) \quad \begin{aligned} \mathbb{E}_0[u_{\mathcal{S}}] &= \frac{\mathbb{E}_0[u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \geq \nu\}}] + \mathbb{E}_0[u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}]}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)} \\ &\geq \nu - \frac{\mathbb{E}_0[(-u_{\mathcal{T}}) \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}]}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)}. \end{aligned}$$

We start with the numerator and with a change of measure we have

$$(D.8) \quad \mathbb{E}_0[-u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}] = \mathbb{E}_{\infty}[e^{u_{\mathcal{T}}} (-u_{\mathcal{T}}) \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}] \leq \mathbb{E}_{\infty}[-u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}].$$

We can now strengthen this inequality as follows:

$$\begin{aligned}
\mathbb{E}_\infty[-u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}] &= \mathbb{E}_\infty[-u_{\mathcal{T}_0^-} \mathbb{1}_{\{\mathcal{T}_0^- \leq \mathcal{T}_r^+\}}] \leq \mathbb{E}_\infty[-u_{\mathcal{T}_0^-}] \\
&\leq \sup_{r \geq 0} \mathbb{E}_\infty[-u_{\mathcal{T}_r^-} - r] \leq \frac{\mathbb{E}_\infty[(u_1)^2]}{\mathbb{E}_\infty[-u_1]} \\
&\leq \frac{\sum_{k=1}^K \mathbb{E}_\infty[(u_1^k - I_\infty^k)^2] + (\sum_{k=1}^K I_\infty^k)^2}{\sum_{k=1}^K I_\infty^k} \\
&= \frac{\bar{\sigma}_\infty^2}{\bar{I}_\infty} + K \bar{I}_\infty,
\end{aligned}
\tag{D.9}$$

where $\bar{I}_i = \frac{1}{K} \sum_{k=1}^K I_i^k$ is the average, over all sensors, of the Kullback-Leibler information numbers and $\bar{\sigma}_i^2 := \frac{1}{K} \sum_{k=1}^K \text{Var}_i\{u_1^k\}$ the average, over all sensors, of the variances of the local likelihood ratios u_1^k , under the probability measure \mathbb{P}_i , $i = 0, \infty$. The second inequality in the second line in (D.9) follows from Lorden's [12] upper bound for the average overshoot, strengthened by observing that $(u_1^-)^2 \leq (u_1)^2$.

Furthermore, for the denominator in (D.7) we have

$$\begin{aligned}
\mathbb{P}_0(u_{\mathcal{T}} \geq \nu) &= \mathbb{P}_0(T_\nu^+ < T_0^-) \geq \mathbb{P}_0(T_0^- = \infty) = \frac{1}{\mathbb{E}_0[T_0^+]} = \frac{K \bar{I}_0}{\mathbb{E}_0[u_{\mathcal{T}_0^+}]} \\
&\geq \frac{K \bar{I}_0}{\sup_{r \geq 0} \mathbb{E}_0[u_{\mathcal{T}_r^+} - r]} \geq \frac{(K \bar{I}_0)^2}{K \bar{\sigma}_0^2 + (K \bar{I}_0)^2} = \frac{\bar{I}_0^2}{K^{-1} \bar{\sigma}_0^2 + \bar{I}_0^2} \\
&\geq \frac{\bar{I}_0^2}{\bar{\sigma}_0^2 + \bar{I}_0^2}.
\end{aligned}
\tag{D.10}$$

The second equality in the first line is a classical result of random walk theory (see for example Siegmund [37, Corollary 8.39, Page 173]), whereas the third equality in the first line is an application of Wald's identity. The second inequality in the second line is again the upper bound provided by Lorden [12] for the overshoot, while the last inequality is true because $K \geq 1$.

From (D.8), (D.9) and (D.10) we obtain

$$\frac{\mathbb{E}_0[(-u_{\mathcal{T}}) \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}]}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)} \leq \frac{\bar{\sigma}_\infty^2 + K(\bar{I}_\infty)^2}{\bar{I}_\infty} \frac{\bar{\sigma}_0^2 + \bar{I}_0^2}{\bar{I}_0^2} = \Theta(K)$$

and consequently from (D.7) it follows that $\mathbb{E}_0[u_S] \geq \nu - \Theta(K)$. It remains to find a lower bound for γ in terms of ν . From the false alarm constraint and (D.6) we have

$$\gamma = \mathbb{E}_\infty[-u_S] = \frac{\mathbb{E}_\infty[-u_{\mathcal{T}}]}{\mathbb{P}_\infty(u_{\mathcal{T}} \geq \nu)}.
\tag{D.11}$$

For the expectation in the numerator, we can obtain the following upper bound

$$\begin{aligned}
 \mathbb{E}_\infty[-u_{\mathcal{T}}] &= \mathbb{E}_\infty[-u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}] + \mathbb{E}_\infty[-u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \geq \nu\}}] \\
 (D.12) \quad &\leq \mathbb{E}_\infty[-u_{\mathcal{T}} \mathbb{1}_{\{u_{\mathcal{T}} \leq 0\}}] \leq \frac{\bar{\sigma}_\infty^2}{\bar{I}_\infty} + K\bar{I}_\infty,
 \end{aligned}$$

where the final inequality follows from (D.9). In order to obtain a lower bound for the probability $\mathbb{P}_\infty(u_{\mathcal{T}} \geq \nu)$ in the denominator we start with a change of measure, thus

$$(D.13) \quad \mathbb{P}_\infty(u_{\mathcal{T}} \geq \nu) = \mathbb{E}_0[e^{-u_{\mathcal{T}}} \mathbb{1}_{\{u_{\mathcal{T}} \geq \nu\}}] = \mathbb{E}_0[e^{-u_{\mathcal{T}}} | u_{\mathcal{T}} \geq \nu] \mathbb{P}_0(u_{\mathcal{T}} \geq \nu).$$

Then, with an application of the conditional Jensen inequality we have

$$\begin{aligned}
 \mathbb{E}_0[e^{-u_{\mathcal{T}}} | u_{\mathcal{T}} \geq \nu] &\geq \exp(-\mathbb{E}_0[u_{\mathcal{T}} | u_{\mathcal{T}} \geq \nu]) \\
 &\geq \exp\left(-\nu - \frac{\mathbb{E}_0[(u_{\mathcal{T}} - \nu) \mathbb{1}_{\{u_{\mathcal{T}} \geq \nu\}}]}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)}\right) \\
 (D.14) \quad &\geq \exp\left(-\nu - \frac{\sup_{r \geq 0} \mathbb{E}_0[u_{T_r^+} - r]}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)}\right) \\
 &\geq \exp\left(-\nu - \frac{\frac{\bar{\sigma}_0^2}{\bar{I}_0} + K\bar{I}_0}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)}\right).
 \end{aligned}$$

where in the last inequality we have used, again, Lorden's [12] upper bound for the maximal average overshoot. Combining (D.13) and (D.14) we obtain

$$\begin{aligned}
 \mathbb{P}_\infty(u_{\mathcal{T}} \geq \nu) &\geq \exp\left(-\nu - \frac{\frac{\bar{\sigma}_0^2}{\bar{I}_0} + K\bar{I}_0}{\mathbb{P}_0(u_{\mathcal{T}} \geq \nu)}\right) \mathbb{P}_0(u_{\mathcal{T}} \geq \nu) \\
 (D.15) \quad &\geq \exp\left(-\nu - \frac{\frac{\bar{\sigma}_0^2}{\bar{I}_0} + K\bar{I}_0}{\frac{\bar{I}_0^2}{\bar{\sigma}_0^2 + \bar{I}_0^2}}\right) \frac{\bar{I}_0^2}{\bar{\sigma}_0^2 + \bar{I}_0^2},
 \end{aligned}$$

where the second inequality follows from (D.10). Then, from (D.11), (D.12) and (D.15) we have

$$\gamma \leq \left(\frac{\bar{\sigma}_\infty^2}{\bar{I}_\infty} + K\bar{I}_\infty\right) \exp\left(\nu + \frac{\frac{\bar{\sigma}_0^2}{\bar{I}_0} + K\bar{I}_0}{\frac{\bar{I}_0^2}{\bar{\sigma}_0^2 + \bar{I}_0^2}}\right) \left(\frac{\bar{I}_0^2}{\bar{\sigma}_0^2 + \bar{I}_0^2}\right)^{-1}.$$

Taking logarithms we obtain $\log \gamma \leq \Theta(\log K) + \nu + K\Theta(1)$, which implies that $\log \gamma \leq \nu + \Theta(K)$ and completes the proof. \square

LEMMA 7. *For the expectation of $\mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}]$ we have the following upper bound*

$$(D.16) \quad \mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}] \leq \log \gamma + K\Theta(\Delta)$$

PROOF. Let us first of all define:

$$(D.17) \quad M := \max_{1 \leq k \leq K} \max_{1 \leq j \leq d^k} \{\bar{\Lambda}_j^k, \underline{\Lambda}_j^k\},$$

It is clear that KM is an upper bound for the overshoot $\tilde{y}_{\tilde{\mathcal{S}}} - \tilde{\nu}$, thus

$$(D.18) \quad \tilde{u}_{\tilde{\mathcal{S}}} \leq \tilde{y}_{\tilde{\mathcal{S}}} \leq \tilde{\nu} + KM \leq \log \gamma - \log(\bar{I}_\infty) + KM,$$

where the first inequality is due to the fact that $\inf_{0 \leq s \leq t} \tilde{u}_s \leq 0$, whereas the third inequality is due to Lemma 4. The quantity $\log(\bar{I}_\infty)$ is independent of the thresholds $\{\bar{\Delta}^k, \underline{\Delta}^k\}$ and remains bounded as $K \rightarrow \infty$, thus $\log(\bar{I}_\infty) = \Theta(1)$. Therefore, it remains to show that $M = \Theta(\Delta)$. Indeed, for each $1 \leq j \leq d^k$ and $1 \leq k \leq K$ we have:

$$(D.19) \quad \bar{\Delta}^k \leq \bar{\Delta}_j^k \leq \bar{\Lambda}_j^k = \bar{\Delta}_j^k + \bar{R}_j^k.$$

From the proof of Lemma 2 it is clear that each \bar{R}_j^k is an $O(1)$ term as $\Delta \rightarrow \infty$ (and $o(1)$ as $d^k \rightarrow \infty$), thus $\bar{\Lambda}_j^k$ is bounded above and below by a $\Theta(\Delta)$ term, which proves that $\bar{\Lambda}_j^k = \Theta(\Delta)$. Similarly it can be shown that $\underline{\Lambda}_j^k = \Theta(\Delta)$ and this completes the proof. \square

LEMMA 8.

$$(D.20) \quad \mathbb{E}_0[u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}] \leq K\Theta(\Delta) + \theta \mathbb{E}_0[m_{\tilde{\mathcal{S}}}],$$

where $\theta := \max_{1 \leq k \leq K} \theta^k$.

PROOF. We first observe that for any t and k we have

$$(D.21) \quad \begin{aligned} u_t^k - \tilde{u}_t^k &= u_t^k - u_{\tau_{m_t^k}^k}^k + u_{\tau_{m_t^k}^k}^k - \tilde{u}_{\tau_{m_t^k}^k}^k \\ &\leq \bar{\Delta}^k + \sum_{n=1}^{m_t^k} [\ell_n^k - \tilde{\ell}_n^k] \leq \bar{\Delta}^k + \sum_{n=1}^{m_t^k} \eta_n^k. \end{aligned}$$

The first inequality is due to the fact that $u_t^k - \tilde{u}_t^k$ is upper bounded by $\bar{\Delta}^k$ between transmissions, whereas the second inequality follows from (5.9).

If we now replace t with $\tilde{\mathcal{S}}$, take expectation with respect to \mathbf{P}_0 and apply Lemma 5 (in particular, (D.3)), we obtain

$$(D.22) \quad \mathbb{E}_0[u_{\tilde{\mathcal{S}}}^k - \tilde{u}_{\tilde{\mathcal{S}}}^k] \leq \bar{\Delta}^k + (\mathbb{E}_0[m_{\tilde{\mathcal{S}}}^k] + 1) \mathbb{E}_0[\eta_n^k] \leq \bar{\Delta}^k + \theta^k + \theta^k \mathbb{E}_0[m_{\tilde{\mathcal{S}}}^k],$$

where the second inequality follows from Lemma 2. Then, summing over k we obtain

$$(D.23) \quad \mathbb{E}_0[u_{\tilde{\mathcal{S}}} - \tilde{u}_{\tilde{\mathcal{S}}}] \leq \sum_{k=1}^K \mathbb{E}_0[u_{\tilde{\mathcal{S}}}^k - \tilde{u}_{\tilde{\mathcal{S}}}^k] \leq \sum_{k=1}^K (\bar{\Delta}^k + \theta^k) + \sum_{k=1}^K \theta^k \mathbb{E}_0[m_{\tilde{\mathcal{S}}}^k],$$

which implies (D.20), since $\Delta^k = \Theta(\Delta)$ and $\theta^k = \mathcal{O}(1)$ as $\Delta \rightarrow \infty$. \square

LEMMA 9. *For the average number of messages received by the fusion center up to time $\tilde{\mathcal{S}}$ we have the following bound*

$$(D.24) \quad \mathbb{E}_0[m_{\tilde{\mathcal{S}}}] \leq \frac{\log \gamma}{\Theta(\Delta)} + K \Theta(1).$$

PROOF. For every k we have $\tilde{u}_{\tilde{\mathcal{S}}}^k = \sum_{n=1}^{m_{\tilde{\mathcal{S}}}^k} \tilde{\ell}_n^k$ and $|\tilde{\ell}_n^k| \leq M$ for every n , where M is defined as in (D.17). Therefore, from Lemma 5, in particular (D.4), it follows that

$$(D.25) \quad \mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}^k] \geq \mathbb{E}_0[m_{\tilde{\mathcal{S}}}^k] \mathbb{E}_0[\tilde{\ell}_1^k] - 2M.$$

Then, summing over k we obtain

$$\mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}] \geq \sum_{k=1}^K \left[\mathbb{E}_0[m_{\tilde{\mathcal{S}}}^k] \mathbb{E}_0[\tilde{\ell}_1^k] - 2M \right] \geq \left(\min_{1 \leq k \leq K} \mathbb{E}_0[\tilde{\ell}_1^k] \right) \mathbb{E}_0[m_{\tilde{\mathcal{S}}}] - 2KM$$

and consequently

$$(D.26) \quad \mathbb{E}_0[m_{\tilde{\mathcal{S}}}] \leq \frac{\mathbb{E}_0[\tilde{u}_{\tilde{\mathcal{S}}}] + 2MK}{\min_{1 \leq k \leq K} \mathbb{E}_0[\tilde{\ell}_1^k]} \leq \frac{\log \gamma - \log(\bar{I}_{\infty}) + 3KM}{\min_{1 \leq k \leq K} \mathbb{E}_0[\tilde{\ell}_1^k]},$$

where the second inequality is due to (D.18). Since $M = \Theta(\Delta)$, which was shown in the proof of Lemma 7, it suffices to show that $\mathbb{E}_0[\tilde{\ell}_1^k] \geq \Theta(\Delta)$ for every $1 \leq k \leq K$, since in this case (D.26) gives

$$(D.27) \quad \mathbb{E}_0[m_{\tilde{\mathcal{S}}}] \leq \frac{\log \gamma + \Theta(1) + K \Theta(\Delta)}{\Theta(\Delta)} \leq \frac{\log \gamma}{\Theta(\Delta)} + K \Theta(1).$$

Indeed, since $\bar{\Lambda}_j^k \geq \bar{\Delta}_j^k \geq \bar{\Delta}^k$ and $\underline{\Lambda}_j^k \leq M$, we obtain

$$\begin{aligned}
 \text{(D.28)} \quad \mathbb{E}_0[\tilde{\ell}_1^k] &= \sum_{j=1}^{d^k} \left[\bar{\Lambda}_j^k \mathbb{P}_0(z_1^k = j) - \underline{\Lambda}_j^k \mathbb{P}_0(z_1^k = -j) \right] \\
 &\geq \bar{\Delta}^k \mathbb{P}_0(z_1^k > 0) - M \mathbb{P}_0(z_1^k < 0) \\
 &= \bar{\Delta}^k - (\bar{\Delta}^k + M) \mathbb{P}_0(z_1^k < 0) = \Theta(\Delta) + o(1).
 \end{aligned}$$

The last equality is due to the fact that $M = \Theta(\Delta)$ and

$$\bar{\Delta}^k \mathbb{P}_0(z_1^k < 0) = \bar{\Delta}^k e^{-\Delta^k} \mathbb{E}_\infty[e^{\ell_1^k + \Delta^k} \mathbb{1}_{\{\ell_1^k < -\Delta^k\}}] = \bar{\Delta}^k e^{-\Delta^k} \Theta(1),$$

since from renewal theory it is well-known that $\mathbb{E}_\infty[e^{\ell_1^k + \Delta^k} \mathbb{1}_{\{\ell_1^k < -\Delta^k\}}]$ is an asymptotically convergent, thus bounded, term as $\bar{\Delta}^k, \underline{\Delta}^k \rightarrow \infty$. \square

REFERENCES

- [1] BASSEVILLE, M. AND NIKIFOROV, I. V. (1993) *Detection of Abrupt Changes: Theory and Applications*. NJ Prentice-Hall, Engelwood Cliffs. [MR1210954](#)
- [2] BEIBEL, M. (1996). A note on Ritov's Bayes approach to the minimax property of the CUSUM procedure. *Ann. Stat.* **24**(4) 1804–1812. [MR1416661](#)
- [3] CHRONOPOULOU, A. AND FELLOURIS, G. (2013). Optimal sequential change detection for fractional diffusion-type processes. *J. App. Prob.* **50**(1), to appear.
- [4] CROW, R. W. AND SCHWARTZ, S. C. (1996). Quickest detection for sequential decentralized decision systems. *IEEE Trans. Aerosp. Electron. Syst.* **32** 267–283.
- [5] DAYANIK, S., POOR, H. V. AND SEZER, S. O. (2008). Multisource Bayesian sequential change detection. *Ann. Appl. Probab.* **18**(2) 552–590. [MR2399705](#)
- [6] FELLOURIS, G. AND MOUSTAKIDES, G. V. (2011). Decentralized sequential hypothesis testing using asynchronous communication. *IEEE Trans. Inf. Th.* **57**(1) 534–548. [MR2814070](#)
- [7] Gapeev, P.V. (2005). The disorder problem for compound Poisson processes with exponential jumps. *Ann. Appl. Probab.* **15** 487–499. [MR2115049](#)
- [8] HADJILIADIS, O., ZHANG, H. AND POOR, H. V. (2009). One-shot schemes for decentralized quickest change detection. *IEEE Trans. Inf. Th.* **55**(7) 3346–3359. [MR2598025](#)
- [9] HAWKINS, D. M. AND OLWELL, D. H. (1998). *Cumulative Sum Control Charts and Charting for Quality Improvement*. Springer-Verlag, New York. [MR1601506](#)
- [10] LAI, T. L. (2001). Sequential analysis: Some classical problems and new challenges (with discussion). *Stat. Sinica* **11** 303–408. [MR1844531](#)
- [11] — (1995). Sequential change-point detection in quality control and dynamical systems. *J. Roy. Statist. Soc. Ser. B* **57** 613–658. [MR1354072](#)
- [12] LORDEN, G. (1970). On excess over the boundary. *Ann. Math. Stat.* **41**(2) 520–527. [MR0254981](#)
- [13] — (1971). Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42** 1897–1908. [MR0309251](#)
- [14] MEI, Y. (2005). Information bounds and quickest change detection in decentralized decision systems. *IEEE Tr. Inf. Th.* **51**(7), 2669–2681, [MR2246385](#)

- [15] MOUSTAKIDES, G.V. (1986). Optimal stopping times for detecting changes in distributions. *Ann. Stat.* **14**(4) 1379–1387. [MR0868306](#)
- [16] — (1998). Quickest detection of abrupt changes for a class of random processes. *IEEE Tran. Inf. Th.* **44**(5) 1965–1968 [MR1664071](#)
- [17] — (2004). Optimality of the CUSUM procedure in continuous time. *Ann. Stat.* **32**(1) 302–315. [MR2051009](#)
- [18] — (2006). Decentralized CUSUM change detection. *Proc. 9th IEEE Int. Conf. Inf. Fusion*, Florence, Italy.
- [19] — (2008). Sequential change detection revisited. *Ann. Stat.* **36**(2) 787–807. [MR2396815](#)
- [20] PAGE, E. S. (1954). Continuous inspection schemes. *Biometrika* **41** 100–115. [MR0088850](#)
- [21] PESKIR, G. AND SHIRYAEV, A.N. (2002). Solving the Poisson disorder problem. In *Advances in Finance and Stochastics. Essays in Honour of Dieter Sondermann (K. Sandmann and P. Schnbucher, eds.)* 295–312. Springer, Berlin. [MR1929384](#)
- [22] POLLAK, M. (1985). Optimal detection of a change in distribution. *Ann. Stat.* **13**, 206–227. [MR0773162](#)
- [23] POLUNCHENKO, A. AND TARTAKOVSKY, A.G. (2012). State-of-the-art in sequential change-point detection. *Methodol. Comput. Appl. Probab.* **14**(3), 649–684.
- [24] POOR, H.V. AND HADJILIADIS, O. (2009). *Quickest Detection*. Cambridge University Press, UK. [MR2482527](#)
- [25] RAGHUNATHAN, V., SCHURGERS, C., PARK, S. AND SRIVASTAVA, M.B. (2002). Energy-aware wireless microsensor networks. *IEEE Sig. Proc. Mag.* **19**(2) 40–50.
- [26] SEZER, S.O. (2010). On the Wiener disorder problem. *Ann. Appl. Probab.* **20**(4) 1537–1566. [MR2676947](#)
- [27] SHEWHART, W.A. (1931). *Economic Control of Quality of Manufactured Product*. Van Nostrand, New York.
- [28] A. N. SHIRYAEV (1978). *Optimal Stopping Rules*. Springer, New York. [MR0468067](#)
- [29] — (1996). Minimax optimality of the method of cumulative sums (CUSUM) in the case of continuous time. *Russ. Math. Surv.* **51** 750–751. [MR1422244](#)
- [30] — (2011). Quickest detection problems: Fifty years later. *Seq. Anal.* **294** 345–385. [MR2747531](#)
- [31] TARTAKOVSKY, A.G., ROZOVSKY, B., BLAZEK, R. AND H. KIM (2006). Detection of Intrusion in Information Systems by Sequential Chang-Point Methods. *Stat. Method.* **3** 252–293 [MR2240956](#)
- [32] TARTAKOVSKY A. G. AND VEERAVALLI V.V. (2008). Asymptotically optimal quickest change detection in distributed sensor systems. *Seq. Anal.* **27** 441–475. [MR2460208](#)
- [33] TARTAKOVSKY, A.G., POLLAK, M. AND POLUNCHENKO, A.S. (2011). Third-order Asymptotic Optimality of the Generalized Shiryaev-Roberts Detection Procedures. *Theory Probab. Appl.* **58**(3) 534–565.
- [34] TSITSIKLIS, J. N. (1993). Extremal properties of likelihood-ratio quantizers. *IEEE Trans. Comm.* **41** 550–558.
- [35] VEERAVALLI, V.V. (1999). Sequential decision fusion: Theory and applications. *J. Fran. Inst.* **336** 301–322 [MR1674584](#)
- [36] — (2001). Decentralized quickest change detection. *IEEE Tran. Inf. Th.* **47**(4) 1657–1665. [MR1830119](#)
- [37] SIEGMUND, D. (1985). *Sequential Analysis, Tests and Confidence Intervals*. Springer-Verlag, New York. [MR799155](#)
- [38] YILMAZ, Y., MOUSTAKIDES, G.V. AND WANG, X. (2012). Cooperative sequential spectrum sensing based on event-triggered sampling. *IEEE Trans. Signal. Process.*

60(9) 4509–4524.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF SOUTHERN CALIFORNIA
LOS ANGELES, CA 90089-2532, USA
E-MAIL: fellouri@usc.edu

DEPARTMENT OF ELECTRICAL AND
COMPUTER ENGINEERING
UNIVERSITY OF PATRAS
26500 RION, GREECE
E-MAIL: moustaki@upatras.gr